on V. This gives an equation of 1-chains:

$$S^p(\gamma_1) - S^p(\gamma_1') - \tau_1 \;=\; S^p(\gamma_2') - S^p(\gamma_2) + \tau_2 .$$

The left side is a 1-chain on U, the right side a 1-chain on V, so the 1-chain they define is a 1-chain on $U \cap V$. And the boundary of this chain is

$$\partial(S^p(\gamma_1) - S^p(\gamma_1') - \tau_1) \;=\; \partial\gamma_1 - \partial\gamma_1' - \partial\tau_1 \;=\; \partial\gamma_1 - \partial\gamma_1' ,$$

which shows that $\partial\gamma_1 - \partial\gamma_1'$ is a 0-boundary on $U \cap V$, and completes the proof. □

Lemma 10.3. *The boundary operator* $\partial\colon H_1(U \cup V) \to H_0(U \cap V)$ *is a homomorphism of abelian groups.*

Proof. This follows readily from the definition. For if α is represented by $\gamma_1 + \gamma_2$, and α' is represented by $\gamma_1' + \gamma_2'$, with γ_1 and γ_1' 1-chains on U and γ_2 and γ_2' 1-chains on V, then $\alpha \pm \alpha'$ is represented by $(\gamma_1 \pm \gamma_1') + (\gamma_2 \pm \gamma_2')$, so

$$\partial(\alpha \pm \alpha') \;=\; [\partial(\gamma_1 \pm \gamma_1')] \;=\; [\partial(\gamma_1)] \pm [\partial(\gamma_1')] \;=\; \partial\alpha \pm \partial\alpha' . \quad \square$$

10b. Mayer–Vietoris for Homology

If $U_1 \subset U_2$, we saw in Chapter 6 that there are homomorphisms from H_0U_1 to H_0U_2 and from H_1U_1 to H_1U_2. Given two open sets U and V we therefore have a diagram

$$\begin{tikzcd}
 & H_1U \arrow[dr] & & & H_0U \arrow[dr] & \\
H_1(U\cap V) \arrow[ur] \arrow[dr] & & H_1(U\cup V) \arrow[r, "\partial"] & H_0(U\cap V) \arrow[ur] \arrow[dr] & & H_0(U\cup V) . \\
 & H_1V \arrow[ur] & & & H_0V \arrow[ur] &
\end{tikzcd}$$

The Mayer–Vietoris story gives the relations among all these groups and homomorphisms. This can be described in a series of assertions, moving from right to left in the diagram.

MV(i). *Any element in* $H_0(U \cup V)$ *is the sum of the images of an element in* H_0U *and an element in* H_0V.

Proof. Any 0-chain that represents an element in $H_0(U \cup V)$ can be written as a sum of a 0-chain in U and a 0-chain in V. □

MV(ii). *An element in H_0U and an element in H_0V have the same image in $H_0(U \cup V)$ if and only if they come from some element in $H_0(U \cap V)$.*

Proof. Let b and c be 0-chains representing elements in H_0U and H_0V. If they have the same image in $H_0(U \cup V)$, there is a 1-chain γ on $U \cup V$ with $\partial\gamma = b - c$. By subdividing the paths in γ to be sufficiently small, so that the image of each subdivided path is contained in U or in V, we may find 1-chains γ_1 on U and γ_2 on V so that $\partial(\gamma_1 + \gamma_2) = \partial\gamma$. Then $b - \partial\gamma_1 = c + \partial\gamma_2$, and the left side of this is a 0-chain on U, and the right side is a 0-chain on V, so $a = b - \partial\gamma_1 = c + \partial\gamma_2$ is a 0-chain on $U \cap V$. Since a is homologous to b on U and to c on V, the class in $H_0(U \cap V)$ represented by a maps to the classes represented by b in H_0U and c in H_0V. The converse is immediate from the definitions. □

MV(iii). *An element in $H_0(U \cap V)$ maps to zero in H_0U and in H_0V if and only if it is the image by ∂ of some element in $H_1(U \cup V)$.*

Proof. We saw during the definition of the boundary ∂ that a boundary always has this form. Conversely, if $\zeta = \partial\gamma_1$ and $\zeta = \partial\gamma_2$ for 1-chains γ_1 on U and γ_2 on V, then $[\zeta] = \partial(\alpha)$, where α is the class represented by the 1-cycle $\gamma_1 - \gamma_2$. □

MV(iv). *An element in $H_1(U \cup V)$ maps to zero in $H_0(U \cap V)$ if and only if it is the sum of an element coming from H_1U and an element coming from H_1V.*

Proof. If $\alpha = [\gamma_1 + \gamma_2]$ for some 1-cycle γ_1 on U and some 1-cycle γ_2 on V, we may take γ_1 and γ_2 for the 1-chains in the definition of the boundary, so $\partial(\alpha) = [\partial\gamma_1] = 0$. Conversely, if α is in the kernel of ∂, and we write $\alpha = [\gamma_1 + \gamma_2]$, with γ_i as in the definition of the boundary, then $\partial\gamma_1 = -\partial\gamma_2$ is a 0-boundary on $U \cap V$, so $\partial\gamma_1 = \partial\tau$ for some 1-chain τ on $U \cap V$. Therefore,

$$\alpha = [(\gamma_1 - \tau) + (\gamma_2 + \tau)],$$

and $\gamma_1 - \tau$ is a 1-cycle on U and $\gamma_2 + \tau$ is a 1-cycle on V. □

MV(v). *An element in H_1U and an element in H_1V have the same image in $H_1(U \cup V)$ if and only if they come from some element in $H_1(U \cap V)$.*

Proof. Let β and γ be 1-cycles on U and V representing the two elements. If they have the same image in $H_1(U \cup V)$, there is an equation

$$\beta - \gamma = \Sigma n_i \partial \Gamma_i$$

for some maps Γ_i from $[0, 1] \times [0, 1]$ to $U \cup V$. We apply the subdividing operator S, which was introduced in the proof of Lemma 10.2, to this equation. As in that lemma, if S is applied sufficiently many times, the images of the subdivided rectangles will be contained in U or in V, and we will have an equation

$$S^p\beta - S^p\gamma = \Sigma n_i S^p(\partial \Gamma_i) = \delta_1 + \delta_2,$$

where δ_1 is a 1-boundary on U and δ_2 is a 1-boundary on V. Lemma 6.4(b) proves that S takes a 1-cycle to a 1-cycle that is homologous to it. So $S^n\beta - \delta_1 = S^p\gamma + \delta_2$ is a 1-cycle on $U \cap V$ that is homologous to β on U and to γ on V, as required. Again, the converse is obvious. □

In the case of open sets in the plane, there is one more assertion that completes the Mayer–Vietoris story.

MV(vi). *If U and V are open subsets of the plane, an element in $H_1(U \cap V)$ is zero if and only if its images in H_1U and in H_1V are zero.*

Proof. If γ is a 1-cycle representing the element in $H_1(U \cap V)$, if its images are zero in H_1U and H_1V, then the winding number $W(\gamma, P)$ is zero for all P not in U and all P not in V. This means that $W(\gamma, P)$ vanishes for all P not in $U \cap V$, and by Theorem 6.11 this implies that γ is homologous to zero on $U \cap V$. □

Exercise 10.4. If $H_1U = 0$ and $H_1V = 0$, show that the kernel of ∂ is zero. If in addition $U \cap V$ is connected, show that $H_1(U \cup V) = 0$.

Exercise 10.5. If U and V are connected, show that the image of ∂ consists of all classes in $H_0(U \cap V)$ of degree zero.

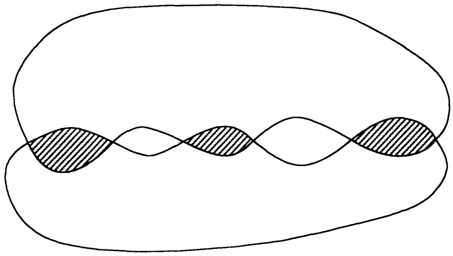
Exercise 10.6. Show that if U and V are connected, and $H_1(U \cup V) = 0$, then $U \cap V$ is also connected.

Exercise 10.7. If $U \cap V$ is connected, and $H_1(U \cap V) = 0$, show that $H_1(U \cup V)$ is isomorphic to the direct sum of H_1U and H_1V.

Exercise 10.8. If the homomorphism from $H_1(U \cap V)$ to H_1V is injective, show that the homomorphism from H_1U to $H_1(U \cup V)$ is also injective.

Exercise 10.9. Suppose A and B are disjoint closed subsets in a space X, and $H_1X = 0$. Show that if $X \setminus A$ and $X \setminus B$ are path-connected, then $X \setminus (A \cup B)$ is also path-connected.

Let us use Mayer–Vietoris to work out an example. We start with a familiar situation, with U and V as simple as possible, which arises when we take the union of two open sets each diffeomorphic to disks.



Suppose U and V are connected, with $H_1U = 0 = H_1V$. Suppose $U \cap V$ has $m \geq 1$ connected components. We claim that $H_1(U \cup V)$ is a free abelian group with $m - 1$ generators. By MV(iv), the boundary map ∂ is injective, and by MV(iii), the image of ∂ consists of elements of $H_0(U \cap V)$ that restrict to zero in $H_0U = \mathbb{Z}$ and $H_0V = \mathbb{Z}$. Now $H_0(U \cap V)$ is the free abelian group on the m connected components of $U \cap V$, so the image of ∂ is the kernel of the homomorphism from this free abelian group to $\mathbb{Z}$ that maps each component to 1. It is easy to see that this kernel is the free abelian group on $m - 1$ generators, which completes the proof. For example, if the components of $U \cap V$ are numbered $W_1, \ldots, W_m$, and e_i is the class that puts coefficient 1 in front of W_i, -1 in front of W_{i+1}, and 0 in front of the other components, then the classes $e_1, \ldots, e_{m-1}$ form a basis for this kernel.

In particular, this recovers the fact that if U is the complement of a point, then $H_1U = \mathbb{Z}$.

Exercise 10.10. Use Mayer–Vietoris to recover the fact that, if U is the complement of n points in the plane, then H_1U is a free abelian group with n generators. Show that small circles around each of the points gives a basis for H_1U.

10c. Variations and Applications

The assertions of the Mayer–Vietoris story can be put in fancy language (but with no change in content) as follows. Define, for $k=0$ and $k=1$, a homomorphism

$$+ \;:\; H_kU \oplus H_kV \;\rightarrow\; H_k(U \cup V)$$

that takes a pair (β, γ) to the sum of the image of β and the image of γ in $H_k(U \cup V)$. In this language, MV(i) says that this homomorphism from $H_0U \oplus H_0V \rightarrow H_0(U \cup V)$ is surjective. Similarly, define a homomorphism

$$- \;:\; H_k(U \cap V) \;\rightarrow\; H_kU \oplus H_kV$$

that sends a class α to the pair $(\beta, -\gamma)$, where β is the image of α in H_kU and γ is the image of α in H_kV. Assertion MV(ii) says that the image of this homomorphism from $H_0(U \cap V)$ to $H_0U \oplus H_0V$ is the kernel of the homomorphism from $H_0U \oplus H_0V$ to $H_0(U \cup V)$.

Given a sequence $\ldots \rightarrow A_{n-1} \rightarrow A_n \rightarrow A_{n+1} \rightarrow \ldots$ of abelian groups and homomorphisms between them, the sequence is called *exact at* A_n if the kernel of the map from A_n to A_{n+1} is equal to the image of the map from A_{n-1} to A_n. The sequence is called *exact* if it is exact at every group in the sequence. For example, the exactness of a sequence $A \rightarrow B \rightarrow 0$ (at B) means precisely that the map from A to B is surjective, and the exactness of $0 \rightarrow A \rightarrow B$ (at A) means that the map from A to B is injective. Assertions MV(i)–MV(vi) can be summarized in the

Theorem 10.11 (Mayer–Vietoris Theorem for Homology). *For any open sets U and V of a topological space, the sequence*

$$H_1(U \cap V) \xrightarrow{-} H_1U \oplus H_1V \xrightarrow{+} H_1(U \cup V) \xrightarrow{\partial} H_0(U \cap V) \xrightarrow{-} H_0U \oplus H_0V \xrightarrow{+} H_0(U \cup V) \rightarrow 0$$

is exact. If U and V are open subsets of the plane, the sequence

$$0 \rightarrow H_1(U \cap V) \xrightarrow{-} H_1U \oplus H_1V \xrightarrow{+} H_1(U \cup V) \xrightarrow{\partial} H_0(U \cap V) \xrightarrow{-} H_0U \oplus H_0V \xrightarrow{+} H_0(U \cup V) \rightarrow 0$$

is exact.

Exercise 10.12. Verify that the exactness of the Mayer–Vietoris sequence at each term is indeed equivalent to the assertions MV(i)–MV(vi).

Exercise 10.13. Suppose A and B are disjoint closed subsets in the plane. Use Mayer–Vietoris to construct an isomorphism

$$H_1(\mathbb{R}^2 \setminus (A \cup B)) \cong H_1(\mathbb{R}^2 \setminus A) \oplus H_1(\mathbb{R}^2 \setminus B).$$

Show that the number of connected components of $\mathbb{R}^2 \setminus (A \cup B)$ is one less than the sum of the numbers of connected components of $\mathbb{R}^2 \setminus A$ and of $\mathbb{R}^2 \setminus B$. Generalize to any finite number of disjoint closed subsets.

Exercise 10.14. Let U be an open subset of the plane, and K a compact subset of U. Show that $H_1(U \setminus K) \cong H_1 U \oplus H_1(\mathbb{R}^2 \setminus K)$. If K has n connected components, recover the isomorphism $H_1(U \setminus K) \cong H_1 U \oplus \mathbb{Z}^n$ of Problem 9.9.

Exercise 10.15. If X is a finite graph (see Problem 5.21) with v vertices and e edges, and X has k connected components, show that $H_1 X$ is a free abelian group with $e - v + k$ generators.

Exercise 10.16. If X is a finite graph in the plane with k connected components, show that $H_1(\mathbb{R}^2 \setminus X)$ is a free abelian group with k generators.

Problem 10.17. For any nonempty space X, we considered in §9a the degree homomorphism deg: $H_0 X \to \mathbb{Z}$ that takes the class of a zero cycle $\Sigma n_i P_i$ to the sum Σn_i of the coefficients. The *reduced* 0th homology group $\tilde{H}_0 X$ of a space X is defined to be the kernel of this map. So X is path-connected if and only if $\tilde{H}_0 X$ is zero. (a) Show that $\tilde{H}_0 X$ is a free abelian group with rank one less that the number of path-connected components of X. (b) Show that, if $U \cap V$ is not empty, there is an exact sequence

$$H_1(U \cap V) \xrightarrow{-} H_1 U \oplus H_1 V \xrightarrow{+} H_1(U \cup V)$$
$$\xrightarrow{\partial} \tilde{H}_0(U \cap V) \xrightarrow{-} \tilde{H}_0 U \oplus \tilde{H}_0 V \xrightarrow{+} \tilde{H}_0(U \cup V) \to 0.$$

This often simplifies computations, since the groups involved are a little smaller.

Problem 10.18. Let U be a connected open set in S^2, and let X be a subset of S^2 homeomorphic to a closed interval such that all of X except for the endpoints is contained in U. Show that $U \setminus X$ is disconnected if and only if the endpoints of X lie in the same connected component of $S^2 \setminus U$.

Problem 10.19. If X is an open set in the plane, and $H_0 X$ and $H_1 X$

have finite ranks, the *Euler characteristic* of X, denoted $\chi(X)$, is defined by

$$\chi(X) \;=\; \operatorname{rank}(H_0X) - \operatorname{rank}(H_1X).$$

Suppose U and V are open sets in the plane, and three of the four open sets U, V, $U \cap V$, and $U \cup V$ have homology of finite ranks. Show that the fourth does also, and that

$$\chi(U \cup V) + \chi(U \cap V) \;=\; \chi(U) + \chi(V).$$

Using the Mayer–Vietoris theorem, homology groups can be used instead of cohomology to prove the Jordan Curve Theorem:

Problem 10.20. With U and V as in the proof of Theorem 5.10, show that the image of the boundary map

$$\partial\colon H_1(U \cup V) = H_1(\mathbb{R}^2 \setminus \{P, Q\}) \;\to\; H_0(U \cap V) = H_0(\mathbb{R}^2 \setminus X)$$

is free with one generator, and deduce that $H_0(\mathbb{R}^2 \setminus X)$ is a free abelian group with two generators, so the complement of X has two connected components. Similarly with U and V as in the proof of Theorem 5.11, show that the image of ∂ is zero, and deduce that $H_0(\mathbb{R}^2 \setminus Y)$ is a free abelian group with one generator, so $\mathbb{R}^2 \setminus Y$ is connected.

The following are some complements to the Jordan Curve Theorem that can be proved by combining Mayer–Vietoris with Corollary 9.4 or Problem 9.9:

Problem 10.21. A compact set K of an open set U is said to *separate* two points if they belong to different connected components of $U \setminus K$. Prove the following version of "Alexander's Lemma": If Y and Z are compact subsets of U, with $Y \cap Z$ connected, and P and Q are two points in $U \setminus (Y \cup Z)$ that are not separated by Y or by Z, then they are not separated by $Y \cup Z$.

Problem 10.22. Let X be a subset of the plane homeomorphic to a circle, P a point on X, and D a disk centered at P. Let A be an open arc of X containing P and contained in D, and let B be the complementary closed arc. Let E be a disk centered at P that doesn't meet B. Show that if two points in E are not separated by X, then they can be connected by a path in $D \setminus X$.

Problem 10.23. Let X be a subset of the plane homeomorphic to a circle, U a connected component of the $\mathbb{R}^2 \setminus X$. Show that, for any

$\varepsilon > 0$, there is a $\delta > 0$ such that any two points of U within distance δ of each other can be connected by a path in the intersection of U with a disk of radius ε. Note that this is false when X is homeomorphic to an interval.

Problem 10.24. Let X be a subset of the plane homeomorphic to a circle, P a point on X, and Q a point not on X. Show that there is a path $\gamma\colon [0,1] \to \mathbb{R}^2$ with $\gamma(0) = Q$, $\gamma(1) = P$, and $\gamma(t) \notin X$ for $0 \le t < 1$.

Problem 10.25. Suppose Y and Z are compact, connected subsets of an open set U, and $Y \cap Z$ has n connected components. (a) Show that $U \setminus (Y \cup Z)$ has at least n connected components. (b) If $U \setminus Y$ and $U \setminus Z$ are connected, show that $U \setminus (Y \cup Z)$ has exactly n connected components.

10d. Mayer–Vietoris for Cohomology

The Mayer–Vietoris story for cohomology is similar to—in fact, dual to—that for homology. If U' is an open subset of a plane open set U, the restriction of 1-forms from U to U' takes closed forms on U to closed forms on U', and exact forms on U to exact forms on U', so we have a linear map of vector spaces

$$H^1(U) \to H^1(U').$$

There is also a linear map $H^0(U) \to H^0(U')$ that takes a locally constant function on U to its restriction to U'. If U and V are open sets in the plane, we therefore have a diagram

$$\begin{tikzcd}[column sep=small]
 & H^0U \arrow[dr] & & & & H^1U \arrow[dr] & \\
H^0(U\cup V) \arrow[ur] \arrow[dr] & & H^0(U\cap V) \arrow[rr, "\delta"] & & H^1(U\cup V) \arrow[ur] \arrow[dr] & & H^1(U\cap V) \\
 & H^0V \arrow[ur] & & & & H^1V \arrow[ur] &
\end{tikzcd}$$

where the maps in each diagram are determined by restrictions, and δ is the coboundary map defined in Chapter 5. The *Mayer–Vietoris story* for De Rham cohomology, for open sets in the plane, is the combination of six assertions:

MV(i). *An element in $H^0(U \cup V)$ maps to zero in H^0U and in H^0V if and only if it is zero.*

MV(ii). *An element in H^0U and an element in H^0V have the same image in $H^0(U \cap V)$ if and only if they come from some element in $H^0(U \cup V)$.*

MV(iii). *An element in $H^0(U \cap V)$ maps to zero in $H^1(U \cup V)$ if and only if it is the difference of an element coming from H^0U and an element coming from H^0V.*

MV(iv). *An element in $H^1(U \cup V)$ maps to zero in H^1U and H^1V if and only if it is the image by δ of an element in $H^0(U \cap V)$.*

MV(v). *An element in H^1U and an element in H^1V have the same image in $H^1(U \cap V)$ if and only if they come from some element in $H^1(U \cup V)$.*

MV(vi). *Any element in $H^1(U \cap V)$ is the difference of an element coming from H^1U and an element coming from H^1V.*

Of these, MV(i) and MV(ii) are straightforward exercises, and MV(iii) and MV(iv) are Propositions 5.7 and 5.9. We prove the nontrivial assertion in MV(v). Given closed 1-forms α on U and β on V, such that there is a function f on $U \cap V$ with $\alpha - \beta = df$ on $U \cap V$, we must construct a closed 1-form ω on $U \cup V$ that differs from α by an exact form on U, and from β by an exact form on V. That is, we want functions f_1 on U and f_2 on V such that

$$\alpha - df_1 = \beta - df_2$$

on $U \cap V$. In other words, we want $df_1 - df_2 = df$. So it will be enough to find f_1 and f_2 so that $f_1 - f_2 = f$ on $U \cap V$. The existence of such f_i follows from Lemma 5.5.

The last case, MV(vi), however is not so obvious. We postpone the proof until Chapter 15, where we make rigorous the "duality" between homology and cohomology. With this duality, in fact, all six assertions for cohomology will be seen to follow from the six assertions for homology. □

The cohomology version of Mayer–Vietoris also has its concise expression as an exact sequence. Define linear maps

$$+ \ : \ H^k(U \cup V) \to H^kU \oplus H^kV$$

by sending a class to the pair consisting of the restrictions of the class to each open set. Define linear maps

$$- \;:\; H^kU \oplus H^kV \;\to\; H^k(U\cap V)$$

by taking a pair (ω_1, ω_2) to the difference $\omega_1|_{U\cap V} - \omega_2|_{U\cap V}$ of its restrictions. As in the case for homology, the assertions MV(i)–MV(vi) are equivalent to the exactness of a sequence:

Theorem 10.26 (Mayer–Vietoris Theorem for Cohomology). *For any open sets U and V in the plane, the sequence*

$$0 \to H^0(U\cup V) \xrightarrow{+} H^0U \oplus H^0V \xrightarrow{-} H^0(U\cap V)$$
$$\xrightarrow{\delta} H^1(U\cup V) \xrightarrow{+} H^1U \oplus H^1V \xrightarrow{-} H^1(U\cap V) \to 0$$

is exact.

It is probably worth remarking that the choice of signs "+" and "−" in both the homology and cohomology versions of Mayer–Vietoris is perfectly arbitrary; those we have chosen will be convenient later.

The following two exercises use Mayer–Vietoris for cohomology, including MV(vi), to strengthen some facts we saw in Chapter 9:

Exercise 10.27. Suppose A and B are disjoint closed subsets in the plane. Use Mayer–Vietoris to construct an isomorphism

$$H^1(\mathbb{R}^2 \setminus (A\cup B)) \;\cong\; H^1(\mathbb{R}^2\setminus A) \oplus H^1(\mathbb{R}^2 \setminus B).$$

Generalize to the complement of any finite number of disjoint closed subsets.

Exercise 10.28. If K is a compact subset of an open set U, and K has n connected components, construct an isomorphism

$$H^1(U\setminus K) \;\cong\; H^1U \oplus \mathbb{R}^n,$$

where the maps to the second factor are given by periods around the components of K.

Exercise 10.29. If U is not empty, define $\widetilde{H}^0U$ to be the quotient space $H^0U/\mathbb{R}$ of the locally constant functions on U by the subspace of constant functions. This is called the *reduced* 0th cohomology group. So $\dim(\widetilde{H}^0U) = \dim(H^0U) - 1$, and U is connected exactly when $\widetilde{H}^0U$ is zero. Show that, if $U\cap V$ is not empty, there is an exact sequence

$$0 \to \tilde{H}^0(U \cup V) \xrightarrow{+} \tilde{H}^0 U \oplus \tilde{H}^0 V \xrightarrow{-} \tilde{H}^0(U \cap V)$$
$$\xrightarrow{\delta} H^1(U \cup V) \xrightarrow{+} H^1 U \oplus H^1 V \xrightarrow{-} H^1(U \cap V) \to 0\,.$$

The result in the following problem was proved by Brouwer:

Problem 10.30. Let K be any compact connected subset of $\mathbb{R}^2$, U a connected component of $\mathbb{R}^2 \setminus K$. Show that the boundary ∂U of U is connected. Is the same true if K is only closed and connected in $\mathbb{R}^2$?

PART VI

COVERING SPACES AND FUNDAMENTAL GROUPS, I

In Chapter 11 we introduce the notion of covering maps, which are generalizations of the polar coordinate mapping, and study their basic properties. Facts about lifting paths and homotopies will generalize what we saw for this special case, which amounted to the basic properties of winding numbers. Many coverings, including the polar coordinate mapping, are examples of G-coverings, arising from an action of a group G on a space, and we emphasize those that arise this way.

We have studied closed paths, and seen that homotopic paths have similar properties. In Chapter 12 we formally introduce the fundamental group, which is the set of homotopy equivalence classes of closed paths starting and ending at a fixed point, the equivalence given by homotopy. In the last section we see how it is related to the first homology group.

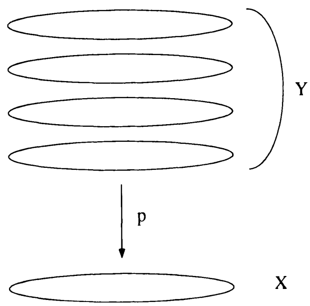
CHAPTER 11

Covering Spaces

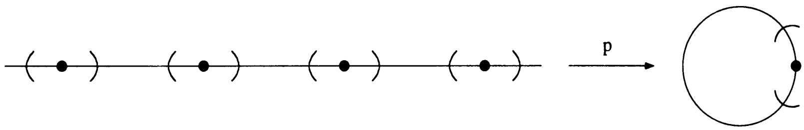
11a. Definitions

If X and Y are topological spaces, a *covering map* is a continuous mapping $p: Y \to X$ with the property that each point of X has an open neighborhood N such that $p^{-1}(N)$ is a disjoint union of open sets, each of which is mapped homeomorphically by p onto N. (If N is connected, these must be the components of $p^{-1}(N)$.) One says that p is *evenly covered* over such N. Such a covering map is called a *covering of* X.

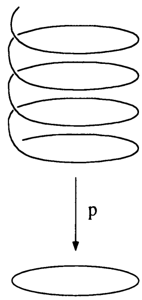
An *isomorphism* between coverings $p: Y \to X$ and $p': Y' \to X$ is a homeomorphism $\varphi: Y \to Y'$ such that $p' \circ \varphi = p$. A covering is called *trivial* if, in the definition, one may take N to be all of X. Equivalently, a covering is trivial if it is isomorphic to the projection of a product $X \times T$ onto X, where T is any set with the discrete topology (all points are closed). So any covering is locally trivial.



The first nontrivial covering is the mapping p: $\mathbb{R}\to S^1$ given by $p(\vartheta) = (\cos(2\pi\vartheta), \sin(2\pi\vartheta))$:



or in a vertical picture:



We saw in Chapter 2 that the *polar coordinate* mapping

$$p : \{(r,\vartheta)\in\mathbb{R}^2 : r>0\} \to \mathbb{R}^2\setminus\{(0,0)\}$$

given by $p(r,\vartheta) = (r\cos(\vartheta), r\sin(\vartheta))$, is a covering. Another is the mapping p_n: $S^1\to S^1$, for any integer $n\geq 1$, given by

$$(\cos(2\pi\vartheta), \sin(2\pi\vartheta)) \mapsto (\cos(2\pi n\vartheta), \sin(2\pi n\vartheta)),$$

or, using complex numbers, by $z\to z^n$; this can be visualized by joining the loose ends in the last picture.

Exercise 11.1. Show that the following are covering maps, where $\mathbb{C}^* = \mathbb{C}\setminus\{0\}$ is the group of nonzero complex numbers: (i) the nth power mapping $\mathbb{C}^*\to\mathbb{C}^*$, $z\to z^n$; and (ii) the exponential mapping $\exp$: $\mathbb{C}\to\mathbb{C}^*$.

Exercise 11.2. Let p: $Y\to X$ be a covering. If X' is any subspace of X, verify that the restriction $Y' = p^{-1}(X')\to X'$ is a covering.

Exercise 11.3. If p: $Y\to X$ is a covering of an open subset X in the plane, show that Y can be given the structure of a differentiable surface in such a way that p is a local diffeomorphism. The components V of $p^{-1}(N)$, for N a connected open set in X over which the covering is trivial, together with the homeomorphisms of N with V given by the inverse of p, give charts covering Y. Similarly, any covering of any manifold has a natural manifold structure.

Exercise 11.4. Show that if X is connected, all fibers of a covering $Y \to X$ have the same cardinality.

When each $p^{-1}(x)$ has cardinality a finite number n, the covering is called an *n-sheeted* covering. It should be emphasized, however, that one cannot distinguish n different "sheets," unless the covering is trivial.

If $p\colon Y \to X$ is a covering, and $f\colon Z \to X$ is a continuous mapping, a continuous mapping $\tilde{f}\colon Z \to Y$ such that $p \circ \tilde{f} = f$ is called a *lifting* of f.

$$\begin{tikzcd} & Y \arrow[d, "p"] \\ Z \arrow[ur, dashed, "\tilde{f}"] \arrow[r, "f"'] & X \end{tikzcd}$$

In the next section we will discuss the question of whether such liftings exist. The following lemma discusses uniqueness: if Z is connected, a lifting is determined by where it maps any one point.

Lemma 11.5. *Let $p\colon Y \to X$ be a covering, and let Z be a connected topological space. Suppose $\tilde{f}_1$ and $\tilde{f}_2$ are continuous mappings from Z to Y such that $p \circ \tilde{f}_1 = p \circ \tilde{f}_2$. If $\tilde{f}_1(z) = \tilde{f}_2(z)$ for one point z in Z, then $\tilde{f}_1 = \tilde{f}_2$.*

Proof. It suffices to show that the set in Z where the mappings agree is open, and its complement where they disagree is also open. If w is in the set where they agree, take a neighborhood N of $p \circ \tilde{f}_i(w)$ that is evenly covered by p. Let $p^{-1}(N)$ be a disjoint union of open sets N_α, with each N_α, mapped homeomorphically to N by p. By continuity, $\tilde{f}_1$ and $\tilde{f}_2$ must map a neighborhood V of w into the same N_α, and since $p \circ \tilde{f}_1 = p \circ \tilde{f}_2$, $\tilde{f}_1$ and $\tilde{f}_2$ must agree on V. Similarly, if w is in the set where the mappings do not agree, $\tilde{f}_1$ and $\tilde{f}_2$ must map a neighborhood V of w into two different (and hence disjoint) N_α's, so they disagree on V. □

11b. Lifting Paths and Homotopies

The work we did in studying the winding number will be abstracted and formalized in the following propositions. In §2a we saw that defining an angle function amounted to lifting a closed path γ in $\mathbb{R}^2 \setminus \{0\}$ to a path $\tilde{\gamma}$ in the right half plane so that the composite of $\tilde{\gamma}$ with the polar coordinate covering p is the given path γ; the difference in second coordinates of $\tilde{\gamma}$ from start to end is then the change in angle along γ. We will now see that this is a general property of covering spaces.

Proposition 11.6 (Path Lifting). *Let $p\colon Y \to X$ be a covering, and let $\gamma\colon [a,b] \to X$ be a continuous path in X. Let y be a point of Y with $p(y) = \gamma(a)$. Then there is a unique continuous path $\tilde{\gamma}\colon [a,b] \to Y$ such that $\tilde{\gamma}(a) = y$ and $p \circ \tilde{\gamma}(t) = \gamma(t)$ for all t in the interval $[a,b]$.*

Proof. The uniqueness comes from Lemma 11.5. When the covering is trivial, the proposition is obvious: there is a unique component of Y that contains y and maps homeomorphically to the component of X that contains $\gamma([a,b])$, and one must simply lift the path to that component using the inverse homeomorphism. For the general case, we apply the Lebesgue lemma (Appendix A) to the open sets $\gamma^{-1}(N)$, where N varies over open sets in X that are evenly covered by p. This gives a subdivision $a = t_0 \leq \ldots \leq t_n = b$ such that each $\gamma([t_{i-1}, t_i])$ is contained in some open set that is evenly covered by p. By the trivial case, there is a lifting of the restriction of γ to $[t_0, t_1]$, giving a path in Y from y to some point y_1. Similarly, there is a lifting of the restriction of γ to $[t_1, t_2]$ that starts at y_1; and one proceeds in n steps until one has lifted the whole path. □

It follows in particular that the final point $\tilde{\gamma}(b)$ of the lifting is determined by γ and by the initial point y. We denote this point by $y*\gamma$, so

$$y*\gamma \;=\; \tilde{\gamma}(b).$$

If this is applied to the polar coordinate covering, and $y = (r_0, \vartheta_0)$ and $y*\gamma = (r_1, \vartheta_1)$ are the initial and final points of the lifting of a path γ, then the difference $\vartheta_1 - \vartheta_0$ is the total change of angle along γ.

Exercise 11.7. Verify this last statement.

One consequence of the path-lifting proposition is the fact that any covering of an interval must be trivial.

The fact that homotopic paths in $\mathbb{R}^2 \setminus \{0\}$ with the same endpoints have the same total change of angle around zero is equivalent to the fact that their liftings are homotopic. This too is a general fact about coverings.

Proposition 11.8 (Homotopy Lifting). *Let $p: Y \to X$ be a covering, and let H be a homotopy of paths in X, i.e., $H: [a,b] \times [0,1] \to X$ is a continuous mapping. Let $\gamma_0(t) = H(t,0)$, $a \le t \le b$, be the initial path. Suppose $\tilde{\gamma}_0$ is a lifting of γ_0. Then there is a unique lifting $\tilde{H}$ of H whose initial path is $\tilde{\gamma}_0$, i.e., $\tilde{H}: [a,b] \times [0,1] \to Y$ is continuous, with $p \circ \tilde{H} = H$ and $\tilde{H}(t,0) = \tilde{\gamma}_0(t)$, $a \le t \le b$.*

Proof. The proof is very much the same as for Proposition 11.6. First apply the Lebesgue lemma to know that there are subdivisions $a = t_0 < t_1 < \ldots < t_n = b$ and $0 = s_0 < s_1 < \ldots < s_m = 1$ so that if $R_{i,j}$ is the rectangle $[t_{i-1}, t_i] \times [s_{j-1}, s_j]$, then each $H(R_{i,j})$ is contained in an evenly covered open set. Then the lifting $\tilde{H}$ is constructed over each piece $R_{i,j}$, say first working across the bottom row, lifting the restriction of H to $R_{1,1}, R_{2,1}, \ldots, R_{n,1}$, then doing the same for the next row $R_{1,2}, R_{2,2}, \ldots, R_{n,2}$, and so on until the entire rectangle has been covered. □

If H is a homotopy of paths from x to x' in X, i.e., $H(a,s) = x$ and $H(b,s) = x'$ for all $0 \le s \le 1$, and if $\tilde{\gamma}_0$ is a path from y to y', then the lifted homotopy $\tilde{H}$ is a homotopy of paths from y to y'. (The fact that $\tilde{H}$ is constant on the sides of the rectangle is guaranteed by the uniqueness of the lifting of the restriction of H to these sides.) In particular, if H is a homotopy from γ_0 to γ_1, then

$$y * \gamma_0 = y * \gamma_1 .$$

For the polar coordinate mapping, this is the preceding assertion about homotopic paths having the same change of angle.

Exercise 11.9. Use the homotopy lifting proposition to show that homotopic closed paths in $\mathbb{R}^2 \setminus \{0\}$ have the same winding number. (The paths in the homotopy are assumed to be closed, but the endpoints can vary during the homotopy.)

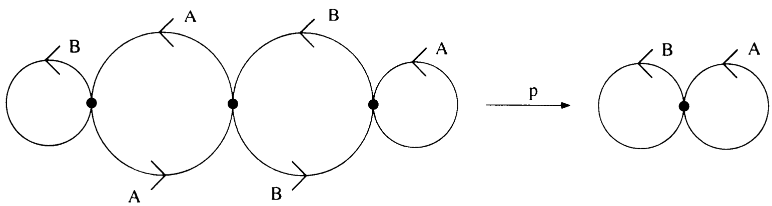
Exercise 11.10. Use the homotopy lifting proposition to prove that any covering of a rectangle (closed or open) must be trivial. Deduce the same for any space homeomorphic to a rectangle, such as a disk.

Exercise 11.11. Suppose $p: Y \to X$ is a covering map, with X a locally connected space (any neighborhood of a point contains a connected neighborhood of the point). (a) Show that X is a union of connected open sets N such that each connected component of $p^{-1}(N)$ is mapped homeomorphically by p onto N. (b) Let Y' be a connected component of Y. Show that image $X' = p(Y')$ is a connected component of X, and the restriction $Y' \to X'$ is a covering map.

Exercise 11.12. Suppose p: $Y \to X$ and p': $Y' \to X$ are covering maps, and φ: $Y \to Y'$ is a continuous map such that $p' \circ \varphi = p$. Suppose X, Y, and Y' are connected, and X is locally connected. Show that φ is a covering map.

Exercise 11.13. If p: $S' \to S$ is an n-sheeted covering, and S is a compact surface, show that S' is also a compact surface. Show, using vector fields and/or triangulations, that the Euler characteristic of S' is n times the Euler characteristic of S. If S and S' are spheres with g and g' handles, show that $g' = ng - n + 1$.

Exercise 11.14. Let X be the space that consist of two circles A and B joined at a point P. Let Y be a space that is two circles and four half-circles, joined as shown, and let p: $Y \to X$ be mapping that takes each piece of Y to the correspondingly labeled piece of X as indicated:



Show that p is a three-sheeted covering. Let γ be the path in X, starting at P, that goes first around the circle A counterclockwise, then around B counterclockwise, then around A clockwise, then around B clockwise. Find the three liftings of γ. Deduce that γ is not homotopic to the constant path at P. Use this to solve Problem 9.14.

11c. G-Coverings

Many important covering spaces arise from the action of a group G on a space Y, with X the space of orbits. Recall that an *action* of a group G on Y (on the left) is a mapping $G \times Y \to Y$, $(g, y) \mapsto g \cdot y$, satisfying:

(1) $g \cdot (h \cdot y) = (g \cdot h) \cdot y$ for all g and h in G and y in Y;
(2) $e \cdot y = y$ for all y in Y, where e is the identity in G; and
(3) the mapping $y \mapsto g \cdot y$ is a homeomorphism of Y for all g in G.

In other words, G defines a group of homeomorphisms of Y. Two points y and y' are in the same *orbit* if there is an element g in G that maps one to the other: $y' = g \cdot y$. Since G is a group, this is an equivalence relation. Let $X = Y/G$ be the set of orbits, or equivalence classes. There is a *projection* p: $Y \to X$ that maps a point to the orbit containing it. The space X is equipped with the quotient topology (i.e., a set U in X is defined to be open when $p^{-1}(U)$ is open in Y).

Exercise 11.15. (a) The group $\mathbb{Z}$ acts on $\mathbb{R}$ by translation: $n \cdot r = r + n$. Show that the quotient $\mathbb{R} \to \mathbb{R}/\mathbb{Z}$ can be identified with the covering map from $\mathbb{R}$ to S^1 described in the first section.

(b) Show that the polar coordinate covering map is the quotient of the right half plane by a $\mathbb{Z}$-action.

(c) The group $G = \mu_n$ of nth roots of unity (which is a cyclic group of order n) in $\mathbb{C}$ acts by multiplication on S^1, regarded as the complex numbers of absolute value 1. Compare the quotient map with the covering p_n: $S^1 \to S^1$ described in §11a.

For another example, the group with two elements acts on the n-sphere with the nontrivial element taking a point to its antipodal point; i.e., $G = \{\pm 1\} \cong \mathbb{Z}/2\mathbb{Z}$ acts on S^n, with $\pm 1 \cdot P = \pm P$. The quotient space is the real projective space $\mathbb{RP}^n$, and the quotient mapping $S^n \to \mathbb{RP}^n$ is a two-sheeted covering.

We say that G acts *evenly*[6] if any point in Y has a neighborhood V such that $g \cdot V$ and $h \cdot V$ are disjoint for any distinct elements g and h in G.

Exercise 11.16. The group μ_n of nth roots of unity acts on $\mathbb{C}$ by

[6] The standard terminology for the notion we are calling "even" is the mouthful "properly discontinuous." The word "discontinuous" does *not* mean that anything is not continuous or otherwise badly behaved; it means that the orbits are discrete subsets of Y. The word "properly" refers to the fact that every compact set only meets finitely many of its translates. If this were not bad enough, the use of "properly discontinuously" in the literature is inconsistent, in that often it is only required that each point have a neighborhood V such that at most finitely many translates $g \cdot V$ of V can intersect. In this case the word "freely" is added, so our "evenly" is then "freely and properly discontinuously."

multiplication. Show that the action is not even, but that the action on the open subset $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ is even.

Lemma 11.17. *If a group G acts evenly on Y, then the projection $p: Y \to Y/G$ is a covering map.*

Proof. The map p is continuous, and open since for V open in Y, the set $p^{-1}(p(V))$ is the union of open sets $g \cdot V$, $g \in G$. If V is taken as in the definition of an even action, then this union is a disjoint union. It is enough to prove that for such V, the mapping from each $g \cdot V$ to $p(V)$ induced by p is a bijection, for then it follows that p is evenly covered over $p(V)$. This is a straightforward verification: since $p(g \cdot y) = p(y)$ for y in V, it is surjective; if $p(g \cdot y_1) = p(g \cdot y_2)$, there is some h in G with $h \cdot g \cdot y_1 = g \cdot y_2$, and the fact that the action is even implies that h is the identity. □

A covering $p: Y \to X$ is called a *G-covering* if it arises in this way from an even action of G on Y. An *isomorphism* of G-coverings is an isomorphism of coverings that commutes with the action of G; i.e., an isomorphism of the G-covering $p: Y \to X$ with the G-covering $p': Y' \to X$ is a homeomorphism $\varphi: Y \to Y'$ such that $p' \circ \varphi = p$ and $\varphi(g \cdot y) = g \cdot \varphi(y)$ for g in G and y in Y. The *trivial* G-covering of X is the product $X \times G \to X$, where G acts by (left) multiplication on the second factor.

Lemma 11.18. *Any G-covering is locally trivial as a G-covering, i.e., if $p: Y \to X$ is a G-covering, then any point in X has a neighborhood N such that the G-covering $p^{-1}(N) \to N$ is isomorphic to the trivial G-covering $N \times G \to N$.*

Proof. In fact, if $N = p(V)$ as in the proof of Lemma 11.17, such a local trivialization is given by

$$p^{-1}(N) \ni g \cdot v \quad \mapsto \quad p(v) \times g \in N \times G. \qquad \square$$

Exercise 11.19. Verify that the actions described in Exercise 11.15 are all even. Show that the action of $\mathbb{Z}^n$ on $\mathbb{R}^n$ by translation is even. Identify the quotient $\mathbb{R}^n/\mathbb{Z}^n$ with the n-dimensional torus $(S^1)^n = S^1 \times \ldots \times S^1$.

Exercise 11.20. Show that any two-sheeted covering has a unique structure of G-covering, where $G = \mathbb{Z}/2\mathbb{Z}$ is the group of order two.

Exercise 11.21. Show that the three-sheeted covering of Exercise 11.14 is not a G-covering.

Exercise 11.22. A *section* of a covering $p: Y \to X$ is a continuous mapping $s: X \to Y$ such that $p \circ s$ is the identity mapping of X. Show that if a G-covering has a section, then the covering is a trivial G-covering.

Exercise 11.23. If $p: Y \to X = Y/G$ is a G-covering that is trivial as a covering, show that it is isomorphic to the trivial G-covering.

Exercise 11.24. Let $p: Y \to X = Y/G$ be a G-covering, and let φ_1 and φ_2 be isomorphisms of G-coverings from Y to Y. If X is connected, and φ_1 and φ_2 agree at one point of Y, show that $\varphi_1 = \varphi_2$.

Exercise 11.25. Let G be the subgroup of the group of homeomorphisms of the plane to itself generated by the translation $(x, y) \mapsto (x+1, y)$ and by the mapping $(x, y) \mapsto (-x, y+1)$. Show that this action of G on $\mathbb{R}^2$ is even, and identify the quotient $\mathbb{R}^2/G$ with the Klein bottle.

Exercise 11.26. Let G be the subgroup of the group of homeomorphisms of the plane to itself generated by the translation $(x, y) \mapsto (x+1, -y)$. Show that this action is even, and identify the quotient with a Moebius band.

Exercise 11.27. If G acts evenly on a space Y, and H is a subgroup of G, show that H also acts evenly. Show that the natural map from Y/H to Y/G is a covering mapping. If n is the index of H in G, this is an n-sheeted covering. Carry this out when G is the group of transformations of the plane from Exercise 11.25, and H is the subgroup generated by the two homeomorphisms $(x, y) \mapsto (x+1, y)$ and $(x, y) \mapsto (x, y+2)$. Identify $\mathbb{R}^2/H$ with a torus, and describe the resulting two-sheeted covering of the Klein bottle.

Exercise 11.28. If a finite group G acts on a Hausdorff space Y, and there are no fixed points (i.e., no y is fixed by any g in G except the identity element), show that the action is even.

Exercise 11.29. If G acts evenly as a group of diffeomorphisms of a differentiable manifold Y, show how to give Y/G the structure of a differentiable manifold in such a way that the projection $Y \to Y/G$ is a local diffeomorphism.

Exercise 11.30. Let $G = \mu_n$ be the group of nth roots of unity, as in Exercise 11.15. The odd-dimensional sphere S^{2m-1} can be realized as

$$S^{2m-1} = \{(z_1, \ldots, z_m) \in \mathbb{C}^m : |z_1|^2 + \ldots + |z_m|^2 = 1\}.$$

The group G acts on S^{2m-1} by $\zeta \cdot (z_1, \ldots, z_m) = (\zeta z_1, \ldots, \zeta z_m)$. Show that this action is even. When n is prime, the quotient space S^{2m-1}/μ_n is a manifold called a *Lens space*.

Exercise 11.31. Suppose $\mathfrak{G}$ is a topological group, i.e., a topological space which is also a group such that the multiplication and inverse maps are continuous. Suppose G is a discrete subgroup of $\mathfrak{G}$, i.e., there is a neighborhood N of the identity e in $\mathfrak{G}$ such that $N \cap G = \{e\}$. Show that left multiplication by G on $\mathfrak{G}$ is an even action. This makes $\mathfrak{G}$ a G-covering of the space $G\backslash\mathfrak{G}$ of right cosets of G in $\mathfrak{G}$. Many of our coverings have this form, e.g., $\mathbb{R}^n \rightarrow (S^1)^n$ is the quotient by the subgroup $\mathbb{Z}^n$; exp: $\mathbb{C} \rightarrow \mathbb{C}^*$ is the quotient of $\mathbb{C}$ by $2\pi i\mathbb{Z}$.

Exercise 11.32. Let $Y = \mathbb{R}^n \setminus \{0\}$, let r be any real number but 0, 1, or -1. Let $G = \mathbb{Z}$ act on Y by $m \cdot v = r^m v$ for $m \in \mathbb{Z}$, $v \in Y$. Show that the action is even, and show that Y/G is homeomorphic to the product $S^1 \times S^{n-1}$.

Problem 11.33 (For those who know about quaternions). The three-sphere S^3 can be identified with the set of unit quaternions:

$$S^3 = \{r + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}: r^2 + x^2 + y^2 + z^2 = 1\},$$

which makes it a topological group. Identify $\mathbb{R}^3$ with the set of pure quaternions $\{x\mathbf{i} + y\mathbf{j} + z\mathbf{k}\}$. For $\mathbf{q} \in S^3$, the mapping $\mathbf{v} \mapsto \mathbf{q} \cdot \mathbf{v} \cdot \mathbf{q}^{-1}$ defines an orthogonal transformation of $\mathbb{R}^3$ with determinant 1, i.e., an element of SO(3). Show that the resulting map $S^3 \rightarrow \mathrm{SO}(3)$ is a surjective homomorphism of groups, with kernel $\{\pm 1\}$. Deduce that this is two-sheeted covering. (This is the *spin* group Spin(3), which, with its generalizations $\mathrm{Spin}(n) \rightarrow \mathrm{SO}(n)$ for all $n \geq 3$, appear frequently in modern mathematics and physics.)

Problem 11.34. Suppose G is a subgroup of the group of distance-preserving transformations of the plane $\mathbb{R}^2$, that satisfies a *uniformity* condition: there is a $d > 0$ such that for all points P in the plane and all g in G other than the identity, $\|g \cdot P - P\| \geq d$.

(a) Show that the action of G on the plane $\mathbb{R}^2$ is even. Let p: $\mathbb{R}^2 \rightarrow \mathbb{R}^2/G = X$ be the resulting G-covering. Define a distance function on X by

$$\mathrm{dist}(Q_1, Q_2) = \mathrm{Min}\{\|P_1 - P_2\|: p(P_1) = Q_1 \text{ and } p(P_2) = Q_2\}.$$

(b) Show that this distance function defines a metric on X.

(c) Prove the analogous result for uniform actions on $\mathbb{R}^3$ (or on any locally compact metric space). This puts a geometric structure on the quotient space X which is locally Euclidean. The reader is invited to look at the resulting "locally Euclidean geometry" on X for some of the groups we have seen. For a general discussion, with a list of possibilities for $\mathbb{R}^2$ and $\mathbb{R}^3$, see Nikulin and Shafarevich (1987).

11d. Covering Transformations

For any covering $p\colon Y \to X$ there is a group $\mathrm{Aut}(Y/X)$ of *covering transformations*, or *deck transformations*:

$$\mathrm{Aut}(Y/X) = \{\varphi\colon Y \to Y : \varphi \text{ is a homeomorphism and } p \circ \varphi = p\}.$$

This is a group by composition of mappings, and it acts on Y in the sense of the preceding section; it is called the *automorphism group* of the covering.

Exercise 11.35. If $Y \to X$ is a trivial n-sheeted covering, and X is connected, show that $\mathrm{Aut}(Y/X)$ is isomorphic to the symmetric group on n letters.

Exercise 11.36. If $p\colon Y \to X$ is the covering of Exercise 11.14, show that $\mathrm{Aut}(Y/X)$ contains only the identity element.

If the covering is a G-covering, there is a canonical homomorphism from G to $\mathrm{Aut}(Y/X)$ that takes g to the homeomorphism $y \mapsto g \cdot y$. By the definition of even action, this homomorphism is injective. It need not be surjective; for example, it is not surjective if the covering is a trivial G-covering, see Exercise 11.35.

Proposition 11.37. *If $p\colon Y \to X$ is a G-covering, and Y is connected, then the canonical homomorphism $G \to \mathrm{Aut}(Y/X)$ is an isomorphism.*

Proof. Fix any point y in Y. Given φ in $\mathrm{Aut}(Y/X)$, since y and $\varphi(y)$ lie in the same orbit, there is a g in G such that $g \cdot y = \varphi(y)$. The covering transformation determined by g and φ both take y to $\varphi(y)$, so by Lemma 11.5 they coincide. $\square$

For a G-covering, G acts transitively on each fiber $p^{-1}(x)$ of p; that is, for y and y' in a fiber, there is a g in G with $g \cdot y = y'$. In addition, this action is faithful: the element g taking y to y' is unique. The following proposition gives a converse to this:

Proposition 11.38. *Let $p\colon Y \to X$ be a covering, with Y connected and X locally connected. Then $Aut(Y/X)$ acts evenly on Y. If $Aut(Y/X)$ acts transitively on a fiber of p, then the covering is a G-covering, with $G = Aut(Y/X)$.*

Proof. We show first that the action is even. For y in Y, let N be a neighborhood of $p(y)$ evenly covered by p, and let V be the neighborhood of y mapping homeomorphically to N by p. If φ and φ' are distinct covering transformations, then $\varphi(V)$ and $\varphi'(V)$ must be disjoint, for if not, then $\varphi^{-1} \circ \varphi'$ has a fixed point in V, and Lemma 11.5 implies that $\varphi^{-1} \circ \varphi'$ is the identity.

Let $G = \mathrm{Aut}(Y/X)$. The covering p factors into the composite of the projection $Y \to Y/G$ followed by a mapping $\bar{p}\colon Y/G \to X$. It follows easily from the definitions that this $\bar{p}$ is a covering mapping; indeed, any open connected set of X that is evenly covered by p is evenly covered by $\bar{p}$. If x is a point in X such that G acts transitively on $p^{-1}(x)$, then the fiber of $\bar{p}$ over x has only one point. By Exercise 11.4, all fibers of $\bar{p}$ have one point, so $\bar{p}$ is a homeomorphism. It follows that p is a G-covering. □

Problem 11.39. Let $p\colon Y \to X$ be a covering, with Y connected and X path-connected. Let $x \in X$, and set $S = p^{-1}(x)$. Any automorphism of the covering restricts to a permutation of S, and the automorphism is determined by this restriction, so $\mathrm{Aut}(Y/X) \subset \mathrm{Aut}(S)$. Show that

$$\mathrm{Aut}(Y/X) \cong \{\varphi \in \mathrm{Aut}(S)\colon \varphi(z * \sigma) = \varphi(z) * \sigma \quad \text{for all} \quad z \in S \text{ and all closed paths } \sigma \text{ at } x\}.$$

CHAPTER 12

The Fundamental Group

12a. Definitions and Basic Properties

The aim of this section is to make the homotopy equivalence classes of paths that start and end at a fixed point in a space into a group. We will see later how this group tells us about the covering spaces of X, as well as determining the first homology and cohomology groups (when X is an open set in the plane). In this chapter it will be convenient again to have all paths defined on the same interval. So a *path* in a topological space X will be a continuous map $\gamma\colon [0,1]\to X$. We say that γ is a path *from* the point $x=\gamma(0)$ *to* the point $x'=\gamma(1)$. In this chapter a *homotopy* of paths will always fix the endpoints x and x', i.e., $H(0,s)=x$ and $H(1,s)=x'$ for all $0\le s\le 1$.

If σ is a path from a point x to a point x', and τ is a path from x' to a point x'', there is a *product* path, denoted $\sigma\cdot\tau$, which is a path from x to x''. It is the path that first traverses σ, then τ, but it must do so at double speed to complete the trip in the same unit time:

$$\sigma\cdot\tau(t) \;=\; \begin{cases}\sigma(2t), & 0\le t\le 1/2,\\ \tau(2t-1), & 1/2\le t\le 1.\end{cases}$$

If σ is a path from x to x', there is an *inverse* path σ^{-1} from x' to x:

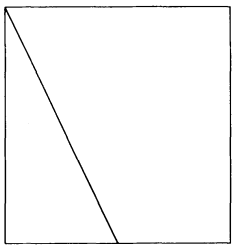
$$\sigma^{-1}(t) \;=\; \sigma(1-t), \qquad 0\le t\le 1.$$

For any point x, let ε_x be the *constant* path at x:

$$\varepsilon_x(t) \;=\; x, \qquad 0\le t\le 1.$$

Exercise 12.1. Show that, for paths from a point x to a point x', the relation of being homotopic is an equivalence relation.

Next we verify that, up to homotopy, these operations satisfy (where defined!) the group axioms. For example, if σ is a path from x to x', there is a homotopy from the path $\varepsilon_x \cdot \sigma$ to σ. This is done by adjusting the time of waiting at x, as indicated in the diagram:



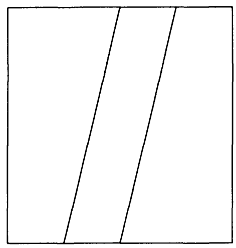
We write down the corresponding homotopy, which is constant on the vertical and slanted lines drawn; on horizontal lines, it is the same as indicated on the top and bottom, but adjusted proportionally:

$$H(t,s) = \begin{cases} x, & 0 \le t \le {}^1\!/\!_2(1-s), \\ \sigma\left(\dfrac{t - {}^1\!/\!_2(1-s)}{1 - {}^1\!/\!_2(1-s)}\right), & {}^1\!/\!_2(1-s) \le t \le 1. \end{cases}$$

Exercise 12.2. Construct a homotopy from $\sigma \cdot \varepsilon_{x'}$ to σ.

In these cases, and those that follow, it is not hard to write down explicit formulas for the homotopy, using a little plane geometry to map rectangles and triangles onto each other, to interpolate between the values we want on boundary lines. It may be worth pointing out that one could also appeal to general results about maps between convex sets that guarantee the existence of such maps; the formulas themselves are not important.

If σ is a path from x to x', τ a path from x' to x'', and μ is a path from x'' to x''', there is a homotopy from $(\sigma \cdot \tau) \cdot \mu$ to $\sigma \cdot (\tau \cdot \mu)$ constructed similarly from the diagram

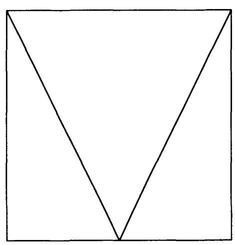


Exercise 12.3. (a) Write down the homotopy indicated by this diagram. (b) Define the path $\sigma \cdot \tau \cdot \mu$ by the formula

$$\sigma \cdot \tau \cdot \mu(t) = \begin{cases} \sigma(3t), & 0 \le t \le 1/3, \\ \tau(3t-1), & 1/3 \le t \le 2/3, \\ \mu(3t-2), & 2/3 \le t \le 1. \end{cases}$$

Show that $\sigma \cdot \tau \cdot \mu$ is homotopic to the paths $(\sigma \cdot \tau) \cdot \mu$ and $\sigma \cdot (\tau \cdot \mu)$.

There is also a homotopy from the path $\sigma \cdot \sigma^{-1}$ to the constant path ε_x. This family of paths does part of the trip specified by σ, rests at the point it has reached, then returns. This is indicated on the following diagram, but note that this time the function is not constant on the diagonal lines:



The homotopy is

$$H(t,s) = \begin{cases} \sigma(2t), & 0 \le t \le 1/2(1-s), \\ \sigma(1-s), & 1/2(1-s) \le t \le 1/2(1+s), \\ \sigma(2-2t), & 1/2(1+s) \le t \le 1. \end{cases}$$

We need one more homotopy: if two paths σ and σ' are homotopic by a homotopy H_1, and τ and τ' are homotopic by a homotopy H_2, then $\sigma \cdot \tau$ and $\sigma' \cdot \tau'$ are homotopic, by the homotopy

$$H(t,s) = \begin{cases} H_1(2t,s), & 0 \le t \le 1/2, \\ H_2(2t-1,s), & 1/2 \le t \le 1. \end{cases}$$

Now, for a point x in X, a *loop at* x is a path that starts and ends at x. Define the *fundamental group of* X *with base point* x, denoted $\pi_1(X,x)$, to be the set of equivalence classes of loops at x, where the equivalence is by homotopy (see Exercise 12.1). We write $[\gamma]$ for the class of the loop γ. The *identity* is the class $e = [\varepsilon_x]$ of the constant path ε_x. Define a product by $[\sigma] \cdot [\tau] = [\sigma \cdot \tau]$. The last displayed homotopy implies that this product is well defined on the equivalence classes. The others imply the equations

$$e \cdot [\sigma] = [\sigma], \qquad [\sigma] \cdot e = e, \qquad [\sigma] \cdot [\sigma^{-1}] = e,$$
$$([\sigma] \cdot [\tau]) \cdot [\mu] = [\sigma] \cdot ([\tau] \cdot [\mu]),$$

in $\pi_1(X,x)$. Since $(\sigma^{-1})^{-1} = \sigma$, the equation $[\sigma^{-1}] \cdot [\sigma] = e$ follows. So this product makes $\pi_1(X,x)$ into a group. From Exercise 12.3(b) we deduce that the product $([\sigma] \cdot [\tau]) \cdot [\mu]$ is also equal to $[\sigma \cdot \tau \cdot \mu]$.

It should be emphasized that endpoints must be fixed during the homotopies discussed here. Otherwise any loop γ would be homotopic to a constant loop, by the homotopy $H(t,s) = \gamma((1-s)t)$.

If $f\colon X \to Y$ is a continuous function, and $f(x) = y$, then f determines a homomorphism of groups

$$f_*\colon \pi_1(X,x) \to \pi_1(Y,y),$$

that takes $[\sigma]$ to $[f \circ \sigma]$.

Exercise 12.4. Verify that this is well defined and a group homomorphism.

The fundamental group is a "covariant functor, on the category of pointed spaces." This means that, if we also have a map $g\colon Y \to Z$ with $g(y) = z$, so that we also have $g_*\colon \pi_1(Y,y) \to \pi_1(Z,z)$, then $(g \circ f)_* = g_* \circ f_*$, i.e., the diagram

$$\begin{tikzcd} \pi_1(X,x) \arrow[rr, "f_*"] \arrow[dr, "(g\circ f)_*"'] & & \pi_1(Y,y) \arrow[dl, "g_*"] \\ & \pi_1(Z,z) & \end{tikzcd}$$

commutes. In addition, if f is the identity, then $f_*: \pi_1(X,x) \to \pi_1(X,x)$ is the identity mapping.

Some of the simple applications of fundamental groups use nothing more than this functoriality. For example, if we know that $\pi_1(X,x)$ is a "complicated" group, and $\pi_1(Y,y)$ is a "simple" group, so that there are no homomorphisms of groups such that the composite $\pi_1(X,x) \to \pi_1(Y,x) \to \pi_1(X,x)$ is the identity, then there can be no continuous maps $f: X \to Y$ and $g: Y \to X$ such that $g \circ f$ is the identity map on X. For example, if $\pi_1(X,x)$ is not the trivial group, and $\pi_1(Y,y)$ is trivial, then X cannot be embedded in Y as a retract.

Exercise 12.5. Show that $\pi_1(D^2,x) = \{e\}$ for any x in the disk, and that the winding number determines an isomorphism of $\pi_1(S^1,(1,0))$ with $\mathbb{Z}$. Deduce that S^1 is not a retract of D^2.

Exercise 12.6. Show that if $X \subset \mathbb{R}^n$ is a subspace that is starshaped about the point x, then $\pi_1(X,x) = \{e\}$.

Although the definition of fundamental group depends on the choice of base point, one gets the same group (up to isomorphism) if one chooses another base point, at least if X is path-connected, or if the two points can be connected by a path. Suppose τ is a path from x to x'. Define a map

$$\tau_\#: \pi_1(X,x) \to \pi_1(X,x')$$

by $[\gamma] \mapsto [\tau^{-1} \cdot (\gamma \cdot \tau)] = [(\tau^{-1} \cdot \gamma) \cdot \tau]$. As before, this is well defined, and is a homomorphism of groups. It is an isomorphism, since $(\tau^{-1})_\#$ gives the inverse homomorphism from $\pi_1(X,x')$ to $\pi_1(X,x)$.

Exercise 12.7. (a) Verify these assertions. (b) If τ' is another path from x to x', show that

$$(\tau')_\#[\gamma] = [\rho]^{-1} \cdot (\tau_\#[\gamma]) \cdot [\rho],$$

where ρ is the loop $\tau^{-1} \cdot \tau'$. In particular, if τ and τ' are homotopic paths from x to x', they determine the same isomorphism on fundamental groups. In general, the displayed equation means that the isomorphism from $\pi_1(X,x)$ to $\pi_1(X,x')$ depends on the choice of path from x to x' only up to inner automorphism.

For this reason, if X is a path-connected space, one often speaks of "the fundamental group" of X, without referring to a base point. This will cause no confusion as long as we are only interested in the group up to isomorphism.

The fundamental group of a ball D^n or of $\mathbb{R}^n$ is trivial, say by Exercise 12.6. We have seen that the circle has an infinite cyclic fundamental group. Let's look next at the n-sphere S^n, for $n \geq 2$. This is almost as simple as it seems. The complement of any point in S^n is homeomorphic to $\mathbb{R}^n$, so any loop that misses any point is homotopic to the constant loop. Although it is possible for a continuous loop to map onto the n-sphere, for any such n, this is not a serious obstruction. We will see a more general reason later, but for now it can be seen directly:

Exercise 12.8. Show that any path in S^n, $n \geq 2$, is homotopic to a path whose image contains no neighborhood of any of its points. Deduce that the fundamental group of S^n is trivial if $n \geq 2$.

Exercise 12.9. Show that the fundamental group of a Cartesian product is the product of fundamental groups of the spaces:

$$\pi_1(X \times Y, x \times y) \cong \pi_1(X, x) \times \pi_1(Y, y).$$

Problem 12.10: If $\mathfrak{G}$ is a topological group, show that $\pi_1(\mathfrak{G}, e)$ is commutative.

12b. Homotopy

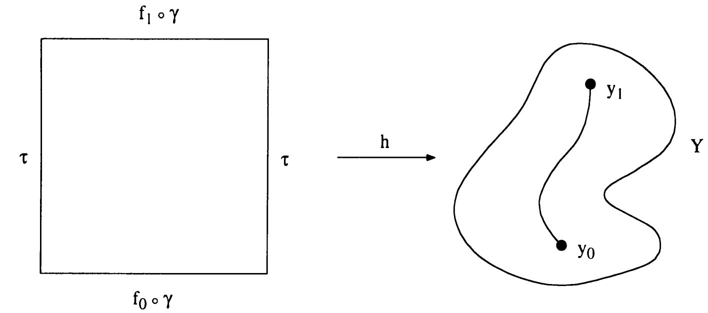
Two continuous maps $f_0: X \to Y$ and $f_1: X \to Y$ are *homotopic* if there is a continuous mapping $H: X \times [0, 1] \to Y$ such that $H(z, 0) = f_0(z)$ and $H(z, 1) = f_1(z)$ for all points z in X; H is a *homotopy from* f_0 *to* f_1. We want to say that homotopic maps determined the same homomorphism on fundamental groups, but to make this precise we must keep track of base points. Let x be a base point in X, and let $y_0 = f_0(x)$ and $y_1 = f_1(x)$. The mapping $\tau(t) = H(x, t)$ is a path from y_0 to y_1.

Proposition 12.11. *The diagram*

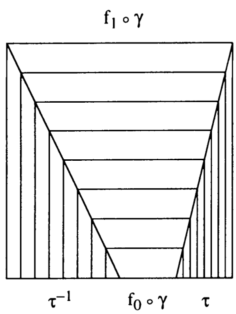
$$\begin{tikzcd} & \pi_1(Y, y_0) \arrow[d, "\tau_\#"] \\ \pi_1(X, x) \arrow[ur, "(f_0)_*"] \arrow[dr, "(f_1)_*"'] & \\ & \pi_1(Y, y_1) \end{tikzcd}$$

commutes, i.e., $\tau_\# \circ (f_0)_* = (f_1)_*$.

Proof. Let γ be a loop at x. We need to construct a homotopy from the path $\tau^{-1}\cdot((f_0\circ\gamma)\cdot\tau)$ to the path $f_1\circ\gamma$. Consider the homotopy h from $[0, 1]\times[0, 1]$ to Y given by $h(t, s) = H(\gamma(t), s)$:



This provides a kind of homotopy between the path $\tau^{-1}\cdot((f_0\circ\gamma)\cdot\tau)$ around the two sides and the bottom of the square and the path $f_1\circ\gamma$ that goes across the top. □



Exercise 12.12. Turn this into a homotopy from $\tau^{-1}((f_0\circ\gamma)\cdot\tau)$ to $f_1\circ\gamma$ by constructing a continuous map from the square to itself that is the identity on the top side, is constant on the two sides, and maps the bottom side to the three sides, as indicated.

As a special case, we have the

Corollary 12.13. *If $f_0(x) = f_1(x) = y$, and H is a homotopy from f_0 to f_1 such that $H(x, s) = y$ for all s, then*

$$(f_0)_* = (f_1)_* \;:\; \pi_1(X, x) \to \pi_1(Y, y).$$

Two spaces X and Y are said to *have the same homotopy type* if there are continuous maps $f\colon X\to Y$ and $g\colon Y\to X$ such that $g\circ f$ is homotopic to the identity map of X and $f \circ g$ is homotopic to the identity map of Y. The map f is called a *homotopy equivalence* if there is such a g.

Exercise 12.14. (a) Show that having the same homotopy type is an equivalence relation. (b) Show that a homotopy equivalence $f: X \to Y$ determines an isomorphism $f_*: \pi_1(X, x) \to \pi_1(Y, f(x))$ of fundamental groups. In particular, if $i: X \to Y$ embeds X as a deformation retract of Y, then $i_*: \pi_1(X, x) \to \pi_1(Y, i(x))$ is an isomorphism.

Exercise 12.15. Classify the following spaces according to homotopy type: (i) a point; (ii) a closed disk; (iii) a circle; (iv) $\mathbb{R}^2$; (v) $\mathbb{R}^n$; (vi) the complement of a point in the plane; (vii) two circles joined at a point; (viii) an annulus; and (ix) the complement of two points in the plane.

Exercise 12.16. Show that the mapping from S^2 to itself that takes (x, y, z) to $(-x, -y, z)$ is homotopic to the identity mapping, and the mapping that takes (x, y, z) to $(x, y, -z)$ is homotopic to the antipodal mapping.

Exercise 12.17. If $f: S^n \to S^n$ is a continuous mapping such that $f(P) \neq P$ for all P, show that f is homotopic to the antipodal mapping. If $f(P) \neq -P$ for all P, show that f is homotopic to the identity mapping.

Problem 12.18. If n is odd, show that the identity mapping on S^n is homotopic to the antipodal mapping.

It is a general fact that for even n, the antipodal map on S^n is not homotopic to the identity map. We will prove this in Chapter 23.

Problem 12.19. An orthogonal $(n+1) \times (n+1)$ matrix determines a mapping from S^n to itself. Show that two such mappings are homotopic if and only if they have the same determinant. If n is even, show that such a mapping is homotopic to the identity if the determinant is 1, and to the antipodal mapping if the determinant is -1.

Problem 12.20. Compute the fundamental group of the space $\mathrm{GL}_2^+(\mathbb{R})$ of (2×2)-matrices with positive determinant, which gets its topology as an open subspace of $\mathbb{R}^4$.

12c. Fundamental Group and Homology

For any topological space X, with base point x in X, there is a homomorphism from the fundamental group $\pi_1(X,x)$ to the homology group H_1X, that takes the class $[\gamma]$ of a loop γ at x to the homology class of γ, regarded as a closed path or 1-chain. It takes the constant path ε_x at x to 0. The fact that it is well defined amounts to the fact that homotopic paths define the same homology class, as we saw in Lemma 6.4. The same lemma showed that the homology class of a product $\sigma \cdot \tau$ of loops is the sum of the homology classes of the loops, which shows that the mapping is a homomorphism of groups.

Exercise 12.21. If $f\colon X \to Y$ is a continuous mapping, and $f(x) = y$, show that the following diagram commutes:

$$\begin{array}{ccc} \pi_1(X,x) & \xrightarrow{f_*} & \pi_1(Y,y) \\ \downarrow & & \downarrow \\ H_1X & \xrightarrow[f_*]{} & H_1Y. \end{array}$$

Since H_1X is an abelian group, this homomorphism must vanish on all commutators $a \cdot b \cdot a^{-1}b^{-1}$ in $\pi_1(X,x)$, so it must vanish on the *commutator subgroup* $[\pi_1(X,x), \pi_1(X,x)]$ that consists of all finite products of commutators. This is a normal subgroup of $\pi_1(X,x)$, and the quotient group is sometimes called the *abelianized fundamental group* of X, and denoted $\pi_1(X,x)_{\text{abel}}$. So we have a homomorphism

$$\pi_1(X,x)_{\text{abel}} = \pi_1(X,x)/[\pi_1(X,x), \pi_1(X,x)] \to H_1X.$$

Since the fundamental group $\pi_1(X,x)$ depends only on the path-connected component of X that contains x, we cannot expect the fundamental group to determine the homology group for disconnected spaces. But except for this, the fundamental group determines the homology group:

Proposition 12.22. *If X is a path-connected space, then the canonical homomorphism from $\pi_1(X,x)_{abel}$ to H_1X is an isomorphism.*

Proof. We must define a homomorphism from the abelian group Z_1X of 1-cycles to $\pi_1(X,x)_{\text{abel}}$, and show that the 1-boundaries B_1X map to zero. This will give a map back from H_1X to $\pi_1(X,x)_{\text{abel}}$. To define the map, let $\gamma = \Sigma_i n_i \gamma_i$ be a 1-cycle, with paths γ_i going from points $a(i)$ to $b(i)$ in X. For each point c that occurs as an endpoint of any

γ_i, choose a path τ_c from x to c. Let γ_i' be the loop at x defined by

$$\gamma_i' = \tau_{a(i)} \cdot \gamma_i \cdot \tau_{b(i)}^{-1},$$

where we are using the notation of Exercise 12.3(b). Define the map from Z_1X to $\pi_1(X,x)_{\text{abel}}$ by sending $\gamma = \Sigma_i n_i \gamma_i$ to the class of $\Pi_i[\gamma_i']^{n_i}$; note that the order in products is unimportant since the group $\pi_1(X,x)_{\text{abel}}$ is abelian.

We verify first that this is independent of the choice of paths τ_c. Suppose $\tilde{\tau}_c$ is another path from x to c, for each c. Let $\tilde{\gamma}_i' = \tilde{\tau}_{a(i)} \cdot \gamma_i \cdot \tilde{\tau}_{b(i)}^{-1}$. Let ϑ_c be the loop $\tilde{\tau}_c \cdot \tau_c^{-1}$. Then $[\tilde{\gamma}_i'] = [\vartheta_{a(i)}] \cdot [\gamma_i'] \cdot [\vartheta_{b(i)}^{-1}]$, so

$$\prod_i [\tilde{\gamma}_i']^{n_i} = \prod_i [\gamma_i']^{n_i} \cdot \left(\prod_i [\vartheta_{a(i)}]^{n_i} \cdot \prod_i [\vartheta_{b(i)}]^{-n_i} \right) = \prod_i [\gamma_i']^{n_i},$$

the last equation using the fact that γ is a 1-cycle, so each point c occurs as many times as a starting point $a(i)$ as ending point $b(i)$.

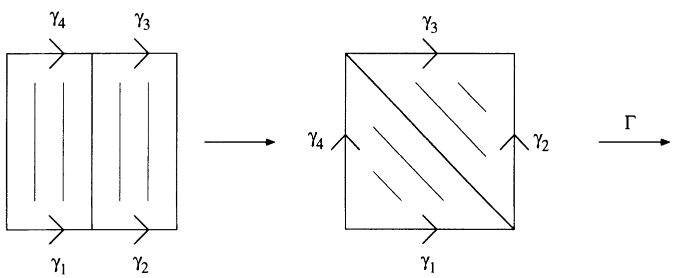
It follows from the definition that this map is a homomorphism from Z_1X to $\pi_1(X,x)_{\text{abel}}$. We next verify that this homomorphism maps boundary cycles to zero. It suffices to show that γ maps to zero when

$$\gamma = \partial\Gamma = \gamma_1 + \gamma_2 - \gamma_3 - \gamma_4,$$

where $\Gamma\colon [0,1] \times [0,1] \to X$ is a continuous mapping, and the boundary is as described in §6a. Let τ_1 and τ_2 be paths from x to the starting and ending points of γ_1, and let τ_3 and τ_4 be paths from x to the starting and ending points of γ_3. Then γ maps to the class of

$$\begin{aligned}
&[\tau_1 \cdot \gamma_1 \cdot \tau_2^{-1}] \cdot [\tau_2 \cdot \gamma_2 \cdot \tau_4^{-1}] \cdot [\tau_4 \cdot \gamma_3^{-1} \cdot \tau_3^{-1}] \cdot [\tau_3 \cdot \gamma_4^{-1} \cdot \tau_1^{-1}] \\
&\qquad = [\tau_1 \cdot \gamma_1 \cdot \gamma_2 \cdot \gamma_3^{-1} \cdot \gamma_4^{-1} \cdot \tau_1^{-1}] = [(\tau_1 \cdot \gamma_1 \cdot \gamma_2) \cdot (\tau_1 \cdot \gamma_4 \cdot \gamma_3)^{-1}].
\end{aligned}$$

To see that this last class is trivial in $\pi_1(X,x)$, it suffices to show that the paths $\tau_1 \cdot \gamma_1 \cdot \gamma_2$ and $\tau_1 \cdot \gamma_4 \cdot \gamma_3$ are homotopic with fixed endpoints. For this it is enough to show that $\gamma_1 \cdot \gamma_2$ and $\gamma_4 \cdot \gamma_3$ are homotopic with fixed endpoints. This is evident from the picture:



An explicit homotopy is given by the formula

$$H(t,s) = \begin{cases} \Gamma(2t(1-s), 2ts), & 0 \leq t \leq {}^1\!/_2, \\ \Gamma((2t-1)s+1-s, (2t-1)(1-s)+s), & {}^1\!/_2 \leq t \leq 1. \end{cases}$$

Finally, we must check that the two homomorphisms we have defined are inverse to each other. It is immediate from the definitions that the composite

$$\pi_1(X,x)_{\mathrm{abel}} \to H_1X \to \pi_1(X,x)_{\mathrm{abel}}$$

is the identity. For the other composite, for a 1-cycle $\gamma = \Sigma_i n_i \gamma_i$, and γ_i' defined as above, the element $\Pi_i[\gamma_i']^{n_i}$ in $\pi_1(X,x)_{\mathrm{abel}}$ maps to

$$\sum_i n_i \gamma_1' = \sum_i n_i(\tau_{a(i)} + \gamma_i - \tau_{b(i)}) = \sum_i n_i \gamma_i = \gamma,$$

using the fact that the boundary of γ is zero. □

So far, all the fundamental groups we have calculated explicitly have been abelian, in which case the proposition says that $\pi_1(X,x) \cong H_1(X)$. We will soon calculate other examples with a nonabelian fundamental group. For now, however, we can appeal to Exercise 11.14 to find an example with a nonabelian fundamental group.

PART VII
COVERING SPACES AND FUNDAMENTAL GROUPS, II

In this part we will see that the two basic notions of covering spaces and fundamental groups are intimately related. In Chapter 13 we see how knowledge of the fundamental group controls the possible coverings a space may have: coverings correspond to subgroups of the fundamental group. There is a universal covering, from which all other coverings can be constructed. In Chapter 14 we use this the other way to prove the Van Kampen theorem, which relates the fundamental group of a union of two spaces to the fundamental groups of the two spaces and the fundamental group of their intersection. This can be regarded as the analogue for the fundamental group of the Mayer–Vietoris theorem for the homology group. The proof we give depends on a correspondence between G-coverings and homomorphisms from the fundamental group to G.

Caution to Beginners. The theorems relating coverings and the fundamental group are stated in their natural generality, with various technical conditions about the spaces involved. This has two advantages: you will have a reasonably complete story when you are finished, and the precise conditions sometimes help in shaping the proofs. However, this also has a large disadvantage: the proliferation of these technical conditions can get in the way of the main ideas. At least for a first reading, it is probably a good idea to simply assume all spaces arising are reasonably behaved—for example, that they are open sets in the plane or a surface or manifold, or perhaps a finite graph. These cases, in fact, will suffice for the applications considered in this text.

CHAPTER 13

The Fundamental Group and Covering Spaces

13a. Fundamental Group and Coverings

We first have the basic:

Proposition 13.1. *If $p\colon Y \to X$ is a covering, and $p(y) = x$, then the induced homomorphism $p_*\colon \pi_1(Y, y) \to \pi_1(X, x)$ is an injection.*

Proof. This is a consequence of the lifting properties of Chapter 11. We must show that the kernel of p_* is $\{e\}$. If σ is a loop at y, and $p_*[\sigma] = e$, there is a homotopy H from $p \circ \sigma$ to the constant path ε_x at x. By homotopy lifting, H lifts to a homotopy $\tilde{H}$ from σ to some path. Since H maps the sides and top of the unit square to the point x, its lifting $\tilde{H}$ (by uniqueness of path-lifting) maps the sides and top of the square to the point y. So $\tilde{H}$ is a homotopy from σ to the constant path ε_y, and $[\sigma] = e$, as required. □

Exercise 13.2. (a) If σ is a loop at x, and $\tilde{\sigma}$ is the unique lifting of σ to a path starting at y, show that $\tilde{\sigma}$ ends at y if and only if its class is in the image of p_*, i.e., $[\sigma] \in p_*(\pi_1(Y, y))$. (b) If σ and σ' are two paths in X from x to x', and $\tilde{\sigma}$ and $\tilde{\sigma}'$ are the lifts to paths in Y starting at y, show that $\tilde{\sigma}$ and $\tilde{\sigma}'$ have the same endpoint if and only if $[\sigma' \cdot \sigma^{-1}]$ is in $p_*(\pi_1(Y, y))$. In the notation of §11b,

$$y * \sigma = y * \sigma' \iff [\sigma' \cdot \sigma^{-1}] \in p_*(\pi_1(Y, y)) .$$

Exercise 13.3. If y' is another point in Y with $p(y') = x$, and y and y' can be connected by a path in Y, show that the image of $\pi_1(Y, y')$ in $\pi_1(X, x)$ is a subgroup conjugate to the image of $\pi_1(Y, y)$. In fact, if σ is a path from y' to y, and $\gamma = p \circ \sigma$ is its image, show that

$$p_*(\pi_1(Y, y')) = [\gamma] \cdot p_*(\pi_1(Y, y)) \cdot [\gamma]^{-1}.$$

Given a covering $p\colon Y \to X$, for any point y with $p(y) = x$, and any loop σ at x, we defined $y * \sigma$ to be the endpoint of the lift of σ that starts at y. This point $y * \sigma$ is also in $p^{-1}(x)$. We saw in §11b that if σ' is a loop homotopic to σ, then $y * \sigma' = y * \sigma$. For any homotopy class $[\sigma]$ in $\pi_1(X, x)$, we can therefore define $y * [\sigma]$ to be $y * \sigma$. This defines a right action of the fundamental group $\pi_1(X, x)$ on the fiber $p^{-1}(x)$:

$$p^{-1}(x) \times \pi_1(X, x) \to p^{-1}(x), \qquad y \times [\sigma] \mapsto y * [\sigma],$$

taking $y \times [\sigma]$ to the endpoint of the lift of σ that starts at y.

Exercise 13.4. Show that this is a right group action; in particular, $y * ([\sigma] \cdot [\tau]) = (y * [\sigma]) * [\tau]$. If Y is path-connected, show that this action is transitive: for any y and y' in $p^{-1}(x)$ there is some $[\sigma]$ with $y * [\sigma] = y'$. Show that the subgroup that acts trivially on a point y is exactly $p_*(\pi_1(Y, y))$. For fixed y, show that this defines a one-to-one correspondence between set of right cosets $\pi_1(X, x)/p_*(\pi_1(Y, y))$ and the fiber $p^{-1}(x)$. In particular, the index of the subgroup $p_*(\pi_1(Y, y))$ in $\pi_1(X, x)$ is the number of sheets of the covering.

We saw in the last chapter that fundamental groups provide an obstruction to the existence of mappings: if there is no map between the groups, there cannot be a map between the spaces. The following proposition shows that, in the case of covering maps, the converse holds: if the fundamental groups say there may be a map, then there will be. Recall that a space X is *locally path-connected* if any neighborhood of any point contains a neighborhood that is path-connected.

Proposition 13.5. *Suppose $p\colon Y \to X$ is a covering, and $f\colon Z \to X$ is a continuous mapping with Z a connected and locally path-connected space. Let $x \in X$, $y \in Y$, $z \in Z$, be points with $p(y) = f(z) = x$. In order for there to be a continuous mapping $\tilde{f}$ $Z \to Y$ with $p \circ \tilde{f} = f$ and $\tilde{H}(z) = y$, it is necessary and sufficient that $f_*(\pi_1(Z, z))$ be contained in $p_*(\pi_1(Y, y))$:*

$$\begin{tikzcd} & Y \arrow[d, "p"] \\ Z \arrow[ur, dashed, "\tilde{f}"] \arrow[r, "f"'] & X \end{tikzcd} \quad \Longleftrightarrow \quad \begin{tikzcd} & \pi_1(Y,y) \arrow[d, hook, "p_*"] \\ \pi_1(Z,z) \arrow[ur, dashed] \arrow[r, "f_*"'] & \pi_1(X,x) \end{tikzcd}$$

Such a lifting $\tilde{f}$, when it exists, is unique.

Proof. The necessity is clear from the functoriality of the fundamental group, and the uniqueness is a special case of Lemma 11.5. For the converse, to construct $\tilde{f}$, given a point w in Z, choose a path γ in Z from z to w, and let $\sigma = f \circ \gamma$, which is a path starting at x in X. Define $\tilde{f}(w)$ to be $y * \sigma$; that is, $\tilde{f}(w)$ is the endpoint of the path that lifts σ and starts at y. We must first show that this is independent of the choice of path. If γ' is another path from z to w, then $f \circ (\gamma' \cdot \gamma^{-1}) = \sigma' \cdot \sigma^{-1}$ is a loop at x. By the hypothesis, $[\sigma' \cdot \sigma^{-1}]$ is in the image of p_*. By Exercise 13.2(b) it follows that $\tilde{\sigma}'$ and $\tilde{\sigma}$ end at the same point.

We must verify that this mapping $\tilde{f}$ is continuous at the point w. Let N be any neighborhood of $f(w)$ that is evenly covered by p, let V be the open set in $p^{-1}(N)$ that maps homeomorphically onto N and that contains $\tilde{f}(w)$, and choose a path-connected neighborhood U of w so that $f(U) \subset N$. We need to show that $\tilde{f}$ maps U into V. For all points w' in U, we may find a path α from w to w' in U, and then we can use $\gamma \cdot \alpha$ as the path from z to w'. The lifting of $f \circ (\gamma \cdot \alpha) = (f \circ \gamma) \cdot (f \circ \alpha)$ is obtained by first lifting $f \circ \gamma$ to $\tilde{\sigma}$, then lifting $f \circ \alpha$. Since the latter lifting stays in V, this shows that $\tilde{f}(U) \subset V$. □

Corollary 13.6. *Let X be a connected and locally path-connected space. Let $p\colon Y \to X$ and $p'\colon Y' \to X$ be two covering maps, with Y and Y' connected and let $p(y) = x$ and $p'(y') = x$. In order for there to be an isomorphism between the coverings preserving the base points, it is necessary and sufficient that*

$$p_*(\pi_1(Y, y)) = p'_*(\pi_1(Y', y')) .$$

Proof. The necessity is clear, and if these subgroups agree, the proposition gives maps $\varphi\colon Y \to Y'$ and $\psi\colon Y' \to Y$ preserving the base points and compatible with projections. (Note that X being locally path-connected implies that covering spaces Y and Y' are also locally path-connected.) Applying Lemma 11.5 to $\psi \circ \varphi$ and $\varphi \circ \psi$ shows that they are isomorphisms. □

Exercise 13.7. Show that two connected coverings of such an X are isomorphic, with an isomorphism that may not preserve base points, if and only if the images of their fundamental groups are conjugate subgroups of the fundamental group of X.

A path-connected space X is called *simply connected* if its fundamental group is the trivial group. Note that this is independent of choice of the base point. In the preceding proposition, if Z is simply connected, it follows that the liftings $\tilde{f}$ always exist.

Corollary 13.8. *A simply connected and locally path-connected space has only trivial coverings.*

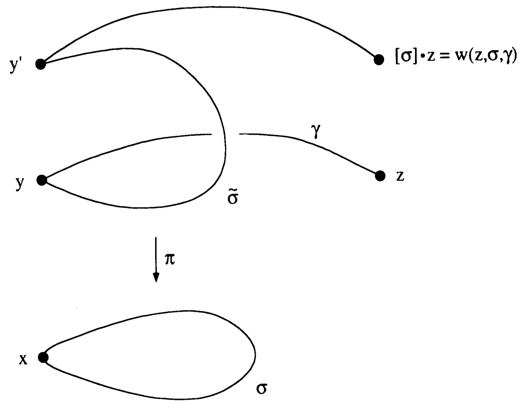
Proof. Suppose $p: Y \to X$ is a covering, with X simply connected. Fix x in X, and for each $y \in p^{-1}(x)$ apply the proposition to get a unique continuous map $s_y: X \to Y$ with $s_y(x) = y$ and $p \circ s_y$ the identity on X. This gives an isomorphism from the trivial covering $X \times p^{-1}(x)$ to Y, by $x \times y \mapsto s_y(x)$. □

Exercise 13.9. Verify that this map is a homeomorphism from the space $X \times p^{-1}(x)$ onto Y.

Problem 13.10. The hypotheses of locally path-connected are needed for the truth of Propositions 13.5 and the two corollaries. *Challenge.* Find an example of a path-connected space X with trivial fundamental group that has a nontrivial connected covering $p: Y \to X$.

13b. Automorphisms of Coverings

Our next goal is to relate the fundamental group of X to the automorphism group of a covering of X. Let $p: Y \to X$ be a covering, with base point y in Y chosen so that $p(y) = x$, and assume Y is path-connected. We want to make $\pi_1(X, x)$ act on the left on Y. Given an element $[\sigma]$ in $\pi_1(X, x)$ and a point z in Y, we therefore want to define a point $[\sigma] \cdot z$ in Y. Let $y' = y * [\sigma]$ be the endpoint of the lift of σ to a path that starts at y. Choose a path γ from y to z in Y. Since $p(y') = p(y) = x$, and $p \circ \gamma$ is a path starting at x, we have a point $y' * (p \circ \gamma)$ that is the endpoint of the lift of the path $p \circ \gamma$ that starts at y'. We want to define $[\sigma] \cdot z$ to be the point $y' * (p \circ \gamma)$. Denote this point $y' * (p \circ \gamma)$ temporarily by $w(z, \sigma, \gamma)$:



Equivalently, $w(z, \sigma, \gamma) = (y * \sigma) * (p \circ \gamma) = y * (\sigma \cdot (p \circ \gamma))$. Suppose, however, that we chose another path γ' from y to z. To have $w(z, \sigma, \gamma') = w(z, \sigma, \gamma)$, we want the lifts of $\sigma \cdot (p \circ \gamma')$ and $\sigma \cdot (p \circ \gamma)$ that start at y to end at the same point. By Exercise 13.2(b), this is the case precisely if the class $[(\sigma \cdot (p \circ \gamma')) \cdot (\sigma \cdot (p \circ \gamma))^{-1}]$ is in $p_*(\pi_1(Y, y))$. Note that $\gamma' \cdot \gamma^{-1}$ is a loop at y, and so

$$\begin{aligned}
[(\sigma \cdot (p \circ \gamma')) \cdot (\sigma \cdot (p \circ \gamma))^{-1}] &= [(\sigma \cdot (p \circ \gamma') \cdot (p \circ (\gamma^{-1})) \cdot \sigma^{-1})] \\
&= [\sigma] \cdot [p \circ (\gamma' \cdot \gamma^{-1})] \cdot [\sigma]^{-1} \\
&= [\sigma] \cdot p_*([\gamma' \cdot \gamma^{-1}]) \cdot [\sigma]^{-1},
\end{aligned}$$

which is an element of $[\sigma] \cdot p_*(\pi_1(Y, y)) \cdot [\sigma]^{-1}$. To know that this is in $p_*(\pi_1(Y, y))$, we need $p_*(\pi_1(Y, y))$ to be a *normal*[7] subgroup of $\pi_1(X, x)$. In this case we see that $w(z, \sigma, \gamma)$ is independent of choice of γ, and depends only on z and the homotopy class $[\sigma]$ of σ.

Assume then that $p_*(\pi_1(Y, y))$ is a normal subgroup of $\pi_1(X, x)$. The above construction determines a mapping

$$\pi_1(X, x) \times Y \to Y, \qquad [\sigma] \times z \mapsto [\sigma] \cdot z,$$

where $[\sigma] \cdot z = w(z, \sigma, \gamma)$. Note that $[\sigma] \cdot z$ is in the same fiber of p as z. Next we want to show that this is a left action of $\pi_1(X, x)$ on Y. The fact that $([\sigma] \cdot [\tau]) \cdot z = [\sigma] \cdot ([\tau] \cdot z)$ follows readily from the definition. For if γ is a path from y to z, then the lift of $\tau \cdot (p \circ \gamma)$, starting at y, is a path from y to $[\tau] \cdot z$. It follows that $[\sigma] \cdot ([\tau] \cdot z)$ is the endpoint of the lift of the path $\sigma \cdot (\tau \cdot (p \circ \gamma))$ that starts at y. The path $(\sigma \cdot \tau) \cdot (p \circ \gamma)$ is homotopic to $\sigma \cdot (\tau \cdot (p \circ \gamma))$, so its lift at y has the same endpoint, and this endpoint is $([\sigma] \cdot [\tau]) \cdot z$. The fact that $[\varepsilon_x] \cdot z = z$

[7] Recall that a subgroup H of a group G is a normal subgroup if $g \cdot H \cdot g^{-1} \subset H$ for all g in G.

follows from the fact that $[\varepsilon_x] \cdot z = (y * \varepsilon_x) * \gamma = y * \gamma$, and $y * \gamma = z$ by definition.

To prove that, for fixed $[\sigma]$, the map $z \mapsto [\sigma] \cdot z$ is continuous, we assume in addition that X is locally path-connected. To see the continuity near a point z, take, as in the preceding proposition, a path-connected neighborhood N of $p(z)$ that is evenly covered by p. Let V and V' be the components of $p^{-1}(N)$ that contain z and $z' = [\sigma] \cdot z$. We must show that $[\sigma] \cdot V$ is contained in V'. For v in V, let α be a path from z to v in V. If γ is a path from y and z, then $\gamma \cdot \alpha$ can be used as the path from y to v, from which it follows that $[\sigma] \cdot v$ is the endpoint of the lift of $p \circ \alpha$ that starts at z'. This lift is in V', which concludes the proof of continuity.

Summarizing, we have constructed a homomorphism from $\pi_1(X, x)$ to the group $\mathrm{Aut}(Y/X)$ of covering transformations. We claim next that this is surjective. Let $\varphi\colon Y \to Y$ be a covering transformation, and suppose $\varphi(y) = y'$. It suffices to find an element $[\sigma]$ in $\pi_1(X, x)$ with $[\sigma] \cdot y = y'$, since, Y being connected, two covering transformations that agree at one point must be identical. Let γ be a path from y to y'. Let $\sigma = p \circ \gamma$. Then $[\sigma] \cdot y$ is the endpoint of the lift of the path σ that starts at y. Since this lift is γ, $[\sigma] \cdot y = y'$, as required.

Finally, we compute the kernel of this homomorphism from $\pi_1(X, x)$ to $\mathrm{Aut}(Y/X)$. As in the preceding step, it suffices to see which $[\sigma]$ act trivially on y, and this happens when the lift of σ at y ends at y, i.e., when $[\sigma]$ is in $p_*(\pi_1(Y, y))$. Putting this all together, we have the:

Theorem 13.11. *Let $p\colon Y \to X$ be a covering, with Y connected and X locally path-connected, and let $p(y) = x$. If $p_*(\pi_1(Y, y))$ is a normal subgroup of $\pi_1(X, x)$, then there is a canonical isomorphism*

$$\pi_1(X, x)/p_*(\pi_1(Y, y)) \xrightarrow{\cong} \mathrm{Aut}(Y/X).$$

The covering is a G-covering, with G being the quotient group $\pi_1(X, x)/p_*(\pi_1(Y, y))$.

The last statement follows from Proposition 11.38. A covering $p\colon Y \to X$ is called *regular* if $p_*(\pi_1(Y, y))$ is a normal subgroup of $\pi_1(X, x)$.

Exercise 13.12. Still assuming Y connected and X locally path-connected, but without assuming $H = p_*(\pi_1(Y, y))$ is a normal subgroup of $\pi_1(Y, y)$, let N be the normalizer of H in $\pi_1(X, x)$, i.e., N is the subgroup of elements g in $\pi_1(X, x)$ such that $g \cdot H \cdot g^{-1} \subset H$. Show that

N acts on the left on Y, and that this determines an isomorphism

$$N/H \cong \mathrm{Aut}(Y/X).$$

Exercise 13.13. Show that the following are equivalent (with Y connected and X locally path-connected): (i) the covering is regular; (ii) the action of $\mathrm{Aut}(Y/X)$ on $p^{-1}(x)$ is transitive; and (iii) for every loop σ at x, if one lifting of σ is closed, then all liftings are closed.

Exercise 13.14. Show that any G-covering $p: Y \to X$, with Y connected and locally path-connected, is a regular covering.

Corollary 13.15. *If $p: Y \to X$ is a covering, with Y simply connected and X locally path-connected, then $\pi_1(X,x) \cong Aut(Y/X)$.*

Corollary 13.16. *If a group G acts evenly on a simply connected and locally path-connected space Y, and $X = Y/G$ is the orbit space, then the fundamental group of X is isomorphic to G.*

In fact, we know by Proposition 11.37 that G is canonically isomorphic to the group of automorphisms of Y over X. Choosing a point y in Y over a point x in X determines an isomorphism of $\pi_1(X,x)$ with $\mathrm{Aut}(Y/X) = G$. □

In general, the isomorphism of G with $\pi_1(X,x)$ depends on the choice of base point y, but only up to inner automorphism. In particular, in case the group is abelian, the isomorphism is independent of choices. For example, this corollary, applied to the mapping from $\mathbb{R}$ to S^1, implies again that $\pi_1(S^1,x) = \mathbb{Z}$.

Applied to the two-sheeted covering $p: S^n \to \mathbb{RP}^n$ from the sphere to the projective space, it follows that

$$\pi_1(\mathbb{RP}^n,x) = \mathbb{Z}/2\mathbb{Z} \qquad \text{for} \quad n \geq 2,$$

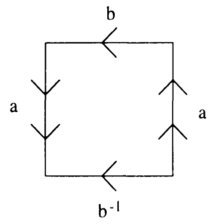
where x is any point in $\mathbb{RP}^n$.

This can be used to give a more conceptual explanation of something we saw in Chapter 4: if $n \geq 2$, there can be no continuous mapping from S^n to S^1 with $g(-P) = -g(P)$ for all P. Such a map would define a continuous mapping $h: \mathbb{RP}^n \to \mathbb{RP}^1$ on quotient spaces:

$$\begin{tikzcd} S^n \arrow[r, "g"] \arrow[d, "p"'] & S^1 \arrow[d, "p'"] \\ \mathbb{RP}^n \arrow[r, "h"'] \arrow[ur, dashed, "\tilde{h}"] & \mathbb{RP}^1, \end{tikzcd}$$

where $p'\colon S^1 \to \mathbb{RP}^1$ is the corresponding mapping for $n = 1$. Since $\pi_1(\mathbb{RP}^1, h(x)) \cong \mathbb{Z}$, the mapping $h_*\colon \pi_1(\mathbb{RP}^n, x) \to \pi_1(\mathbb{RP}^1, h(x))$ must be trivial. Choosing a point y in S^n with $p(y) = x$, by Proposition 13.5 there is a continuous mapping $\tilde{h}$ from $\mathbb{RP}^n$ to S^1 so that $p' \circ \tilde{h} = h$ and $\tilde{h}(x) = g(y)$. Now $\tilde{h} \circ p$ and g are two mappings from S^n to S^1 that map y to $g(y)$, and both, when followed by p', are the map $h \circ p$. By Lemma 11.5, $\tilde{h} \circ p = g$. But $\tilde{h} \circ p$ always takes P and $-P$ to the same point, while g never does. So such g cannot exist.

Exercise 13.17. (a) Compute the fundamental group of the Lens spaces of Exercise 11.30. (b) If the Klein bottle is constructed by identifying sides as shown:



show that the fundamental group has two generators a and b, with one relation $abab^{-1} = e$. In particular, the fundamental group is not abelian. (c) The torus is a two-sheeted covering of the Klein bottle, as in Exercise 11.27. Describe the image of the fundamental group of the torus in the fundamental group of the Klein bottle, and verify that it is a normal subgroup.

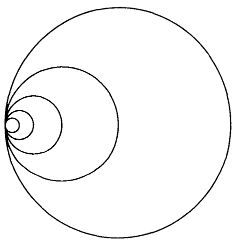
Exercise 13.18. If $p\colon Y \to X$ is the three-sheeted covering of Exercise 11.14, show that $p_*(\pi_1(Y, y))$ is not a normal subgroup of $\pi_1(X, x)$. In particular, $\pi_1(X, x)$ is not an abelian group.

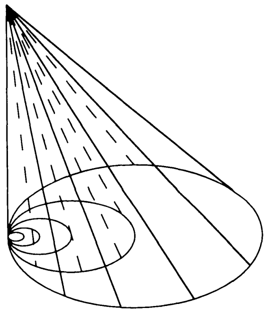
13c. The Universal Covering

In this section we assume that X is a connected and locally path-connected space. A covering $p\colon Y \to X$ is called a *universal covering* if Y is simply connected. It follows from Corollary 13.6 that such a covering, if it exists, is unique, and unique up to canonical isomorphism if base points are specified. As we have seen, the fundamental group of X will be isomorphic to the automorphism group of this covering. The aim of this section is to show that, if one additional property is satisfied, a universal covering always exists.

Suppose we have a universal covering $p\colon Y \to X$. Any point in X has an evenly covered path-connected neighborhood N. Any loop σ in N lifts to a loop $\tilde{\sigma}$ in Y, and, since Y is simply connected, this loop is homotopic in Y to a constant path. It follows that the loop $\sigma = p \circ \tilde{\sigma}$ is homotopic to a constant path in X. A space X is called *semilocally simply connected* if every point has a neighborhood such that every loop in the neighborhood is homotopic in X to a constant path. So being semilocally simply connected is a necessary condition for the existence of a universal covering. Note that if X is *locally simply connected*, i.e., if every neighborhood of a point contains a neighborhood that is simply connected, then X is semilocally simply connected.

The spaces one generally meets, and those we have considered in this book, are all locally simply connected. Any open set in the plane or in $\mathbb{R}^n$, or any manifold, or any finite graph, is locally simply connected. In fact, one has to work a little to produce spaces that are not locally simply connected or semilocally simply connected.





Exercise 13.19. For a positive integer n, let C_n be the circle of radius $1/2^n$ centered at the point $(1/2^n, 0)$. Let $C \subset \mathbb{R}^2$ be the union of all the circles C_n; C is sometimes called a *clamshell*. (a) Show that C is connected and locally path-connected, but not semilocally simply connected. Let X be the cone over C, i.e., $X \subset \mathbb{R}^3$ is the union of all line segments from points in C to the point $(0,0,1)$. (b) Show that X is semilocally simply connected but not locally simply connected.

Suppose we have a universal covering $p\colon Y \to X$, with $p(y) = x$. For any point z in Y, there is a path γ from y to z, which is unique up to homotopy since Y is simply connected. The image $\alpha = p \circ \gamma$ is a path from x to $p(z)$, unique up to homotopy (with fixed endpoints). Conversely, given a path α in X starting at x, it determines a point $z = y * \alpha$ in Y. This identifies Y, at least as a set, with the set of homotopy classes of paths in X that start at a given point x. The idea to the proof of the following theorem is to use this observation in reverse, by using these homotopy classes to construct the universal covering.

Theorem 13.20. *A connected and locally path-connected space X has a universal covering if and only if it is semilocally simply connected.*

Proof. To construct a universal covering, fix x in X and define $\tilde{X}$ to be the set of homotopy classes $[\gamma]$ of paths γ in X that start at x, the homotopies as usual being required to fix endpoints. Assigning to each such class its endpoint defines a function $u: \tilde{X} \to X$. Our task is to put a topology on $\tilde{X}$ so that this is a covering map, and show that $\tilde{X}$ is simply connected.

Call an open set N in X *good* if N is path-connected, and every loop in N at a point z in N is homotopic to the constant path ε_z in X. If γ is a path from x to a point z in N, let $N_{[\gamma]}$ be the subset of $\tilde{X}$ consisting of the homotopy classes $[\gamma \cdot \alpha]$, where α is any path in N that starts at z. Note that $N_{[\gamma]}$ depends only on the homotopy class $[\gamma]$ of γ. Here are some of the properties of these sets:

(1) if β is a path in a good N starting at the endpoint of γ, then $N_{[\gamma \cdot \beta]} = N_{[\gamma]}$;
(2) if a good N is contained in a good N', then $N_{[\gamma]} \subset N'_{[\gamma]}$; and
(3) if γ and γ' are two paths from x to a point z in a good N, then $N_{[\gamma]} = N_{[\gamma']}$ if γ and γ' are homotopic, and $N_{[\gamma]} \cap N_{[\gamma']} = \emptyset$ otherwise. □

Exercise 13.21. Verify these properties.

Now define a subset $\mathcal{O}$ of $\tilde{X}$ to be open if, for any $[\gamma]$ in $\mathcal{O}$, there is a good neighborhood N of the endpoint of γ with $N_{[\gamma]} \subset \mathcal{O}$. It follows from properties (1) and (2) that these open sets form a topology on $\tilde{X}$, and that each of these sets $N_{[\gamma]}$ is open, and the projection from $N_{[\gamma]}$ to N is continuous. In fact, this projection is a homeomorphism, the inverse being given by the map that takes w in N to $[\gamma \cdot \alpha]$, where α is any path in N from the endpoint of γ to w. This is independent of choice of α, since if α' is another, α and α' are homotopic in X, so $\gamma \cdot \alpha$ is homotopic to $\gamma \cdot \alpha'$. This projection is continuous since for smaller good N', N' maps to $N'_{[\gamma]}$.

The map $u: \tilde{X} \to X$ is evenly covered over any good N, since the inverse image of N is a disjoint union of the open sets $N_{[\gamma]}$, where $[\gamma]$ varies over the homotopy classes of paths from x to any given point z of N.

For any path γ starting at x in X, and for s between 0 and 1, let γ_s be the path defined by $\gamma_s(t) = \gamma(st)$, $0 \leq t \leq 1$. The mapping $\tilde{\gamma}$: $[0, 1] \to \tilde{X}$ defined by $\tilde{\gamma}(s) = [\gamma_s]$ is the unique lift of γ to a path in $\tilde{X}$ that starts at the base point $\tilde{x} = [\varepsilon_x]$. In particular, it is a path in $\tilde{X}$ from $\tilde{x}$ to $[\gamma]$, showing that $\tilde{X}$ is connected. Any loop in $\tilde{X}$ at $\tilde{x}$ has the form $\tilde{\gamma}$ for a unique loop γ at x. For $\tilde{\gamma}$ to be a loop, the endpoint $\gamma_1 = \gamma$ must be homotopic to the constant path at x, which implies by the lifting of homotopies that $\tilde{\gamma}$ is homotopic to the constant path at $\tilde{x}$. This shows that $\tilde{X}$ is simply connected, and completes the proof of the theorem. □

Problem 13.22. (a) Suppose X is locally simply connected. Show that, if p: $Y \to X$ and q: $Z \to Y$ are covering maps, then $p \circ q$: $Z \to X$ is also a covering map. (b) Find a counterexample to (a) when X is a clamshell.

13d. Coverings and Subgroups of the Fundamental Group

The theorem of the preceding section will determine correspondence between subgroups of the fundamental group and coverings. For the following proposition, assume that X is connected, locally path-connected, and semilocally simply connected, so that X has a universal covering.

Proposition 13.23. (*a*) *For every subgroup H of $\pi_1(X,x)$ there is a connected covering p_H: $Y_H \to X$, with a base point $y_H \in p_H^{-1}(x)$ so that the image of $\pi_1(Y_H, y_H)$ in $\pi_1(X,x)$ is H. Any other such covering* (*with choice of base point*) *is canonically isomorphic to this one.*

(*b*) *If K is another subgroup of $\pi_1(X,x)$ containing H, there is a unique continuous mapping $p_{H,K}$: $Y_H \to Y_K$ that maps y_H to y_K and is compatible with the projections to X. This mapping is a covering mapping, and if H is a normal subgroup of K, it is a G-covering with $G = K/H$.*

Proof. Let u: $\tilde{X} \to X$ be a universal covering, with $u(\tilde{x}) = x$, and identify $\pi_1(X,x)$ with $\mathrm{Aut}(\tilde{X}/X)$. Any subgroup H of $\pi_1(X,x)$ acts evenly on $\tilde{X}$, and the quotient $\tilde{X} \to \tilde{X}/H = Y_H$ makes $\tilde{X}$ the universal covering of Y_H. The fundamental group of Y_H (with base point y_H the image of $\tilde{x}$) is canonically isomorphic to H. The projection from Y_H to $\tilde{X}/\pi_1(X,x) = X$ is a covering, and the image of its fundamental group in $\pi_1(X,x)$ is H. If H is contained in K, there is a canonical map on the orbit spaces $\tilde{X}/H \to \tilde{X}/K$. The remaining verifications are left to the reader, using the results of the preceding section. $\square$

As seen in the proof, the covering $Y_H \to X$ corresponding to H can be identified with $\tilde{X}/H \to X$, where H acts on $\tilde{X}$ as a subgroup of $\mathrm{Aut}(\tilde{X}/X) = \pi_1(X,x)$. The regular coverings correspond to normal subgroups H. This means (see Exercise 13.14) that every connected G-covering $p: Y \to X$ has the form $\tilde{X}/H \to X$, with

$$\pi_1(X,x)/H \;\cong\; \mathrm{Aut}(Y/X) \;\cong\; G.$$

The correspondence of the proposition is similar to that seen in Galois theory, where subgroups correspond to field extensions, smaller subgroups corresponding to larger extensions. Here smaller subgroups of the fundamental group correspond to larger coverings of the space:

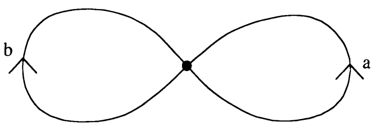
$$\begin{array}{cc}
\tilde{X}, \tilde{x} & \{e\} \\
\downarrow & \cap \\
Y_H, y_H & H \\
\downarrow & \cap \\
Y_K, y_K & K \\
\downarrow & \cap \\
X, x & G
\end{array}$$

If H is a normal subgroup of K, then the covering $Y_H \to Y_K$ is a K/H-covering.

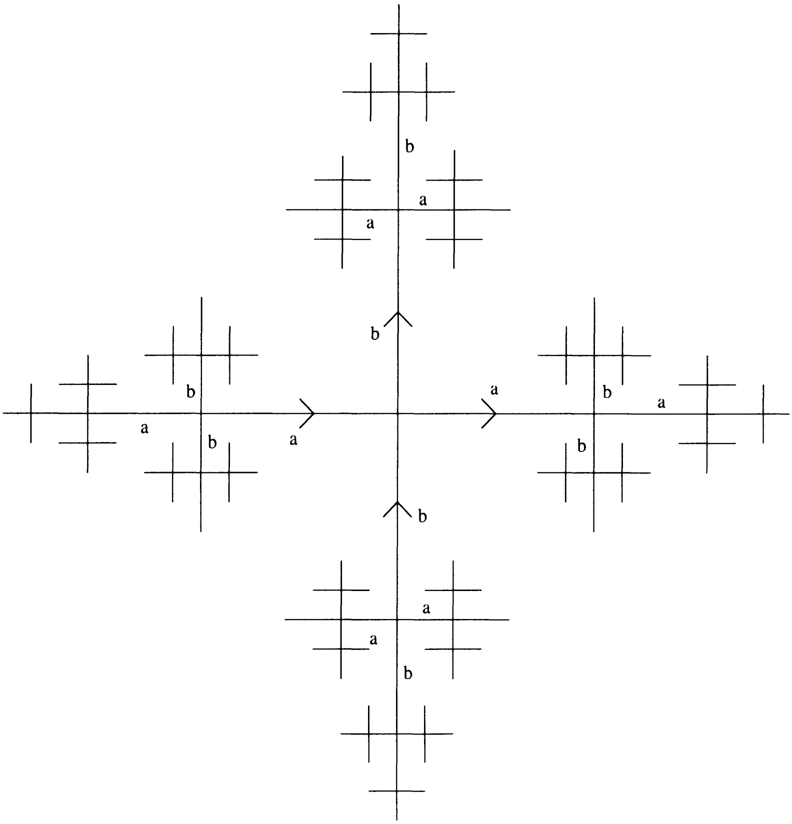
For example, if $X = S^1$ is a circle, the coverings correspond to subgroups of $\pi_1(S^1,(1,0)) = \mathbb{Z}$. The trivial group corresponds to the universal covering $\mathbb{R} \to S^1$, and the subgroups $\mathbb{Z}n \subset \mathbb{Z}$ correspond to the n-sheeted covering $p_n: S^1 \to S^1$ considered in §11a. Up to isomorphism, these are the only connected coverings of a circle.

Exercise 13.24. If $\pi_1(X,x)$ is abelian, show that every connected covering of X is regular.

One simple space whose universal covering we have not yet constructed is the figure 8 space



or the join of two circles at a point. This universal covering can be constructed as an infinite tree. Here part of it is pictured with segments of decreasing size, so that they don't overlap in the plane, but one should imagine that they are all of the same lengths.



Each segment maps to one of the loops of the figure 8 space, the horizontal segments to one and the vertical segments to the other.

Problem 13.25. (a) Show that this space is simply connected, and is the universal covering of the figure 8 space X. (b) If you know what a free group is, show that a free group with two generators acts evenly on this universal covering, with orbit space X. Deduce that the fundamental group of X is this free group. (c) Generalize to the space obtained by joining n circles at a point.

In Chapter 14 we will see a more general method of calculating fundamental groups, which will include another proof of the result of the preceding problem.

The covering corresponding to the commutator subgroup of the fundamental group of X is a covering we may denote by $\widetilde{X}_{\text{abel}}$:

$$\widetilde{X}_{\text{abel}} = \widetilde{X}/[\pi_1(X,x), \pi_1(X,x)] \rightarrow X.$$

This is a G-covering, with $G = \pi_1(X,x)/[\pi_1(X,x), \pi_1(X,x)]$, which is the first homology group H_1X by Proposition 12.22. It is sometimes called the *universal abelian covering* of X. In general a covering is called *abelian* if it is regular with abelian automorphism group.

Exercise 13.26. Show that any connected abelian covering of X has the form $\widetilde{X}_{\text{abel}}/H$, for some subgroup H of H_1X.

We saw in Chapter 6 that for open sets in the plane, the homology class of a closed 1-chain is determined by winding numbers around points not in the set. The next exercise shows that this is true for other nice subsets, but the following problem shows it is not true in general.

Exercise 13.27. Suppose X is a subset of the plane that is contained in some open set U for which there is a retract $r\colon U \rightarrow X$. Show that a closed 1-chain γ on X is homologous to zero if and only if $W(\gamma, P) = 0$ for all P not in X.

Problem 13.28. *Challenge*. Let X be the clamshell of Exercise 13.19. Give an example of a closed path γ on X such that $W(\gamma, P) = 0$ for all P not in X, but such that γ is not a 1-boundary on X.

Exercise 13.29. The universal covering of the complement of the origin in the plane can be realized as the right half plane, via the polar coordinate mapping $(r, \vartheta) \mapsto (r\cos(\vartheta), r\sin(\vartheta))$, and as the entire complex plane $\mathbb{C}$ via the mapping $z \mapsto \exp(z)$. Find an isomorphism between these coverings.

CHAPTER 14

The Van Kampen Theorem

14a. G-Coverings from the Universal Covering

In this section X will denote a connected, locally path-connected, and semilocally simply connected space, so X has a universal covering, denoted $u: \tilde{X} \to X$. All spaces will have base points, and all maps will be assumed to take base points to base points. The base point of X is denoted x, and the base point of $\tilde{X}$ over x is denoted $\tilde{x}$.

We have seen in §13d that for every G-covering $p: Y \to X$, with Y connected, and with base point y, there is a surjective homomorphism of $\pi_1(X,x)$ onto G. If H is the kernel of this homomorphism, so $G \cong \pi_1(X,x)/H$, Y is the quotient of $\tilde{X}$ by the action of H, with y the image of $\tilde{x}$. We want to extend this correspondence to G-coverings that may not be connected. In this case there will only be a homomorphism from $\pi_1(X,x)$ to G, which need not be surjective. Here we will set up this correspondence between G-coverings and homomorphisms directly and rather briefly, omitting some verifications. Other ways to carry this out, with a more general context for these constructions, together with more details about the verifications, are described §16d and §16e.

Suppose $\rho: \pi_1(X,x) \to G$ is a homomorphism from the fundamental group of X to any group G. We will construct from ρ a G-covering $p_\rho: Y_\rho \to X$, together with a base point y_ρ in Y_ρ over x. Give G the discrete topology, so the Cartesian product $\tilde{X} \times G$ is a product of copies of $\tilde{X}$, one for each element in G. The group $\pi_1(X,x)$ acts on the left on $\tilde{X} \times G$ by the rule

$$[\sigma] \cdot (z \times g) = [\sigma] \cdot z \times g \cdot \rho([\sigma]^{-1}) = [\sigma] \cdot z \times g \cdot \rho([\sigma])^{-1},$$

for $[\sigma] \in \pi_1(X,x)$, $z \in \widetilde{X}$, $g \in G$. Here $[\sigma] \cdot z$ is the action of $\pi_1(X,x)$ on $\widetilde{X}$ that was described in §13b, and $g \cdot \rho([\sigma]^{-1})$ is the product in the group G. Define Y_ρ to be the quotient of $\widetilde{X} \times G$ by this action of $\pi_1(X,x)$:

$$Y_\rho = \widetilde{X} \times G/\pi_1(X,x),$$

and let y_ρ be the image of the point $\tilde{x} \times e$ in Y_ρ. Let $\langle z \times g \rangle$ denote the image in Y_ρ of the point $z \times g$ in $\widetilde{X} \times G$. Note that, by the above action of $\pi_1(X,x)$ on $\widetilde{X} \times G$, we have, for z in $\widetilde{X}$, g in G, and $[\sigma]$ in $\pi_1(X,x)$,

$$\langle [\sigma] \cdot z \times g \rangle = \langle z \times g \cdot \rho([\sigma]) \rangle .$$

Define p_ρ: $Y_\rho \to X$ by taking $\langle z \times g \rangle$ to $u(z)$.

The group G acts on Y_ρ by the formula $h \cdot \langle z \times g \rangle = \langle z \times h \cdot g \rangle$, for h and g in G and z in X. (Note that using the right side of G for the left action of $\pi_1(X,x)$ frees up the left side of G for a left action of G!) We claim that this is an even action, making p_ρ: $Y_\rho \to X$ a G-covering. To prove this, let N be any open set in X over which the universal covering u: $\widetilde{X} \to X$ is trivial. By Lemma 11.18 there is an isomorphism of $u^{-1}(N)$ with the product covering $N \times \pi_1(X,x)$, on which $\pi_1(X,x)$ acts on the left on the second factor. This gives homeomorphisms

$$p_\rho^{-1}(N) \cong (N \times \pi_1(X,x)) \times G/\pi_1(X,x) \cong N \times G,$$

the latter homeomorphism by $\langle (u \times [\sigma]) \times g \rangle \mapsto u \times g \cdot \rho([\sigma])$. (The map back takes $u \times g$ to $\langle (u \times e) \times g \rangle$.) These homeomorphisms are compatible with the projections to N, and it follows that, over N, the action of G is even and the covering is a G-covering. Since X is covered by such open sets N, the same is true for the map p_ρ from Y_ρ to X.

Conversely, suppose p: $Y \to X$ is a G-covering, with a base point y over x. From this we construct a homomorphism ρ from $\pi_1(X,x)$ to G. For each $[\sigma]$ in $\pi_1(X,x)$ the element $\rho([\sigma])$ in G is determined by the formula

$$\rho([\sigma]) \cdot y = y * \sigma,$$

where $y * \sigma$ is the endpoint of the lift of the path σ that starts at y. We will need two facts about this operation:

(i) $(z * \sigma) * \tau = z * (\sigma \cdot \tau)$ for $z \in p^{-1}(x)$, σ a loop at x, and τ a path starting at x;

(ii) $g \cdot (z * \gamma) = (g \cdot z) * \gamma$ for $g \in G$, $z \in p^{-1}(x)$, and γ a path starting at x.

The first of these facts is immediate from the definition. The second follows from the fact that if $\tilde{\gamma}$ is lifting of γ starting at z, then the path $t \mapsto g \cdot \tilde{\gamma}(t)$, $0 \leq t \leq 1$, is a lifting of γ that starts at $g \cdot z$. The endpoint of this path, which is $(g \cdot z) * \gamma$ by definition, is $g \cdot \tilde{\gamma}(1)$, and since $\tilde{\gamma}(1) = z * \gamma$, (ii) follows.

We claim now that the ρ defined above is a homomorphism. This is a calculation, using (ii) and (i):

$$\begin{aligned}(\rho([\sigma]) \cdot \rho([\tau])) \cdot y &= \rho([\sigma]) \cdot (\rho([\tau]) \cdot y) = \rho([\sigma]) \cdot (y * \tau) \\ &= (\rho([\sigma]) \cdot y) * \tau = (y * \sigma) * \tau \\ &= y * (\sigma \cdot \tau) = \rho([\sigma] \cdot [\tau]) \cdot y .\end{aligned}$$

Proposition 14.1. *The above constructions determine a one-to-one correspondence between the set of homomorphisms from $\pi_1(X,x)$ to the group G and the set of G-coverings with base point, up to isomorphism*:

$$\mathrm{Hom}(\pi_1(X,x), G) \leftrightarrow \{G\text{-coverings}\}/\text{isomorphism}.$$

Proof. Given a G-covering $p\colon Y \to X$ with base points, from which we constructed a homomorphism ρ, we must now show that the given covering is isomorphic to the covering $p_\rho\colon Y_\rho \to X$ constructed from ρ. To map Y_ρ to Y, we need to map $\tilde{X} \times G$ to Y, and show that orbits by $\pi_1(X,x)$ have the same image. For this we identify the universal covering $\tilde{X}$ as the space of homotopy classes of paths in X starting at x. Define a map

$$\tilde{X} \times G \to Y, \qquad [\gamma] \times g \mapsto g \cdot (y * \gamma) = (g \cdot y) * \gamma .$$

This is easily checked to be continuous. We must check that an equivalent point $([\sigma] \cdot [\gamma]) \times (g \cdot \rho([\sigma])^{-1})$ maps to the same point. By (i) and (ii), this point maps to

$$\begin{aligned}(g \cdot \rho([\sigma])^{-1}) \cdot (y * (\sigma \cdot \gamma)) &= (g \cdot \rho([\sigma])^{-1}) \cdot ((y * \sigma) * \gamma) \\ &= ((g \cdot \rho([\sigma])^{-1}) \cdot (y * \sigma)) * \gamma \\ &= (g \cdot ((y * \sigma) * \sigma^{-1})) * [\gamma] \\ &= (g \cdot (y * (\sigma \cdot \sigma^{-1}))) * [\gamma] = (g \cdot y) * [\gamma] ,\end{aligned}$$

as required. Since the map takes the same values on equivalent points, it gives a mapping from the quotient Y_ρ to Y, which is a mapping of covering spaces of X. This is easily checked to be a mapping of G-coverings, from which it follows that it must be an isomorphism.

Conversely, starting with a homomorphism ρ, we constructed a G-

covering $Y_\rho \to X$, from which we constructed another homomorphism, say $\tilde{\rho}$. We must verify that $\tilde{\rho} = \rho$. Now for $[\sigma]$ in $\pi_1(X,x)$,

$$\begin{aligned}\tilde{\rho}([\sigma]) \cdot \langle \tilde{x} \times e \rangle &= \langle \tilde{x} \times e \rangle * \sigma \ = \ \langle \tilde{x} * \sigma \times e \rangle \\ &= \langle [\sigma] \cdot \tilde{x} \times e \rangle \ = \ \langle \tilde{x} \times e \cdot \rho([\sigma]) \rangle \ = \ \langle \tilde{x} \times \rho([\sigma]) \rangle \\ &= \rho([\sigma]) \cdot \langle \tilde{x} \times e \rangle .\end{aligned}$$

This shows that $\tilde{\rho}([\sigma]) = \rho([\sigma])$, which concludes the proof. □

Exercise 14.2. If $p: Y \to X$ is the G-covering corresponding to a homomorphism $\rho: \pi_1(X,x) \to G$, and X' is a subspace of X that also has a universal covering, with x in X', show that the restriction $p^{-1}(X') \to X'$ of this covering to X' is the G-covering corresponding to the composite homomorphism $\rho \circ i_*$, where $i_*: \pi_1(X',x) \to \pi_1(X,x)$ is induced by the inclusion i of X' in X.

Exercise 14.3. Show that, if base points are ignored, two G-coverings Y_ρ and $Y_{\rho'}$ are isomorphic G-coverings if and only if the homomorphisms ρ and ρ' are *conjugate*, i.e., there is some g in G such that

$$\rho'([\sigma]) \ = \ g \cdot \rho([\sigma]) \cdot g^{-1} \qquad \text{for all} \quad [\sigma] \in \pi_1(X,x).$$

14b. Patching Coverings Together

Suppose X is a union of two open sets U and V. A covering of X restricts to coverings of U and V, which are isomorphic over $U \cap V$. Conversely, suppose we have coverings $p_2: Y_1 \to U$ and $p_2: Y_2 \to V$, and we have an isomorphism of coverings

$$\vartheta: p_1^{-1}(U \cap V) \to p_2^{-1}(U \cap V)$$

of $U \cap V$. Then one may patch (or "glue," or "clutch") these together to get a covering $p: Y \to X$, together with isomorphisms of coverings

$$\varphi_1: Y_1 \xrightarrow{\cong} p^{-1}(U), \qquad \varphi_2: Y_2 \xrightarrow{\cong} p^{-1}(V)$$

of U and of V, so that, over $U \cap V$, $\vartheta = \varphi_2^{-1} \circ \varphi_1$.

One can construct Y as the quotient space of the disjoint union $Y_1 \sqcup Y_2$, by the equivalence relation that identifies a point y_1 in $p_1^{-1}(U \cap V)$ with the point $\vartheta(y_1)$ in $p_2^{-1}(U \cap V)$. (See Appendix A3.) Since ϑ is compatible with maps to X, one gets a mapping p from Y to X. Since the map from Y_1 to Y is a homeomorphism onto its image $p^{-1}U$, which is open in Y, and similarly Y_2 maps homeomorphically onto $p^{-1}V$, one sees that the restriction of p to the inverse image of

U is isomorphic to $Y_1 \to U$, and the restriction over V is isomorphic to $Y_2 \to V$. From this it follows in particular that p is a covering map.

If each of $Y_1 \to U$ and $Y_2 \to V$ is a G-covering, for a fixed group G, and ϑ is an isomorphism of G-coverings, then $Y \to X$ gets a unique structure of a G-covering in such a way that the maps from Y_1 and Y_2 commute with the action of G.

Occasionally the following generalization is useful. Suppose we have a collection X_α of open sets, $\alpha \in \mathscr{A}$, whose union is X, and a collection p_α: $Y_\alpha \to X_\alpha$ of covering maps. Suppose, for each α and β, we have an isomorphism

$$\vartheta_{\beta\alpha}\colon\ p_\alpha^{-1}(X_\alpha \cap X_\beta) \to p_\beta^{-1}(X_\alpha \cap X_\beta)$$

of coverings of $X_\alpha \cap X_\beta$. Assume these are compatible, i.e.,

(1) $\vartheta_{\alpha\alpha}$ is the identity on Y_α; and
(2) $\vartheta_{\gamma\alpha} = \vartheta_{\gamma\beta} \circ \vartheta_{\beta\alpha}$ on $p_\alpha^{-1}(X_\alpha \cap X_\beta \cap X_\gamma)$ for all α, β, $\gamma \in \mathscr{A}$.

Then one can patch these coverings together to obtain a covering p: $Y \to X$. One has isomorphisms φ_α: $Y_\alpha \to p^{-1}(X_\alpha)$ of coverings of X_α, such that $\vartheta_{\beta\alpha} = \varphi_\beta^{-1} \circ \varphi_\alpha$ on $p_\alpha^{-1}(X_\alpha \cap X_\beta)$. In addition, the space Y is the union of the open sets $\varphi_\alpha(Y_\alpha)$.

One constructs Y as the quotient space $\sqcup_{\alpha \in \mathscr{A}} Y_\alpha / R$ of the disjoint union of the Y_α by the equivalence relation determined by the $\vartheta_{\beta\alpha}$'s. The assertions about Y and the φ_α are general facts about patching spaces together, as proved in Appendix A3. The map p is determined by the equations $p \circ \varphi_\alpha = p_\alpha$ on Y_α. Since φ_α is a homeomorphism of Y_α onto $p^{-1}(X_\alpha)$, it follows that p is a covering map.

If each p_α: $Y_\alpha \to X_\alpha$ is a G-covering, with fixed G, and each $\theta_{\beta\alpha}$ is an isomorphism of G-coverings, then there is a unique action of G on Y so that each φ_α commutes with the action of G, i.e., $\varphi_\alpha(g \cdot y_\alpha) = g \cdot \varphi_\alpha(y_\alpha)$ for g in G and y_α in Y_α. This gives the patched covering p: $Y \to X$ the structure of a G-covering, so that each φ_α is an isomorphism of G-coverings.

14c. The Van Kampen Theorem

The Van Kampen theorem describes the fundamental group of a union of two spaces in terms of the fundamental group of each and of their intersection, under suitable hypotheses. Let X be a space that is a union of two open subspaces U and V. Assume that each of the spaces U, V and their intersection $U \cap V$ is path-connected, and let x be a point in the intersection. Assume also that all these spaces X, U, V, and $U \cap V$ have universal covering spaces; this is the case, for example, if X is locally simply connected. We have a commutative diagram of homomorphisms of fundamental groups:

$$\begin{tikzcd} & \pi_1(U,x) \arrow[dr, "j_1"] & \\ \pi_1(U \cap V, x) \arrow[ur, "i_1"] \arrow[dr, "i_2"'] & & \pi_1(X, x) . \\ & \pi_1(V, x) \arrow[ur, "j_2"'] & \end{tikzcd}$$

The maps are induced by the inclusions of subspaces, and commutativity means that $j_1 \circ i_1 = j_2 \circ i_2$.

We will describe how $\pi_1(X,x)$ is determined by the other groups (and the above maps between them). The description will not be direct, but will be by a *universal property*. Note that any homomorphism h from $\pi_1(X,x)$ to a group G determines a pair of homomorphisms $h_1 = h \circ j_1$ from $\pi_1(U,x)$ to G and $h_2 = h \circ j_2$ from $\pi_1(V,x)$ to G; the two homomorphisms $h_1 \circ i_1$ and $h_2 \circ i_2$ from $\pi_1(U \cap V, x)$ to G determined by these are the same. The Van Kampen theorem says that $\pi_1(X,x)$ is the "universal" group with this property.

$$\begin{tikzcd} & \pi_1(U,x) \arrow[dr, "j_1"'] \arrow[drr, "h_1"] & & \\ \pi_1(U \cap V, x) \arrow[ur, "i_1"] \arrow[dr, "i_2"'] & & \pi_1(X, x) \arrow[r, dashed, "h"] & G. \\ & \pi_1(V, x) \arrow[ur, "j_2"] \arrow[urr, "h_2"'] & & \end{tikzcd}$$

Theorem 14.4 (Seifert–van Kampen). *For any homomorphisms*

$$h_1\colon \pi_1(U,x) \to G \quad \textit{and} \quad h_2\colon \pi_1(V,x) \to G,$$

such that $h_1 \circ i_1 = h_2 \circ i_2$, *there is a unique homomorphism*

$$h\colon \pi_1(X,x) \to G,$$

such that $h \circ j_1 = h_1$ *and* $h \circ j_2 = h_2$.

Exercise 14.5. Show that $\pi_1(X,x)$, together with the homomorphisms

j_1 and j_2, is determined up to canonical isomorphism by the universal property.

Exercise 14.6. Use the universal property to show that $\pi_1(X,x)$ is generated by the images of $\pi_1(U,x)$ and $\pi_1(V,x)$. Can you prove this assertion directly?

A version of the Van Kampen theorem was found first by Seifert, and the theorem is also known as the Seifert–Van Kampen theorem. The version given here, via universal properties, was given by Fox, see Crowell and Fox (1963). The usual proof of the Van Kampen theorem (without the hypotheses that the spaces all have universal coverings) is rather technical, and for it we refer to Crowell and Fox (1963) or Massey (1991). Here we will give a quick proof, due to Grothendieck (see Godbillon (1971)), using the correspondence between homomorphisms from fundamental groups to a group G and G-coverings. The assumptions assure that each of the spaces X, U, V, and $U \cap V$ has a universal covering space, and that homomorphisms from their fundamental groups to a group G correspond to G-coverings.

In particular, the homomorphisms h_1 and h_2 determine G-coverings $Y_1 \to U$ and $Y_2 \to V$, together with base points y_1 and y_2 over x. The fact that $h_1 \circ i_1$ is equal to $h_2 \circ i_2$ means that the restrictions of these coverings to $U \cap V$ are isomorphic G-coverings, and since $U \cap V$ is connected, there is a unique isomorphism between these G-coverings that maps the base point y_1 to the base point y_2. (The uniqueness is a special case of Exercise 11.24.) By the construction of the preceding section, these two coverings patch together, using this isomorphism over the intersection. This gives a G-covering $Y \to X$ that restricts to the two given G-coverings (and has the same base point). This G-covering corresponds to a homomorphism h from $\pi_1(X,x)$ to G, and the fact that the restricted coverings agree means precisely that $h \circ j_1 = h_1$ and $h \circ j_2 = h_2$. □

Corollary 14.7. *If U and V are simply connected, then X is simply connected.*

Note the important hypothesis in all these theorems, that all spaces, including the intersection $U \cap V$, are connected. It does not apply to the annulus, written as a union of two sets homeomorphic to disks!

Exercise 14.8. If V is simply connected, show that j_1: $\pi_1(U,x) \to \pi_1(X,x)$

is surjective, with kernel the smallest normal subgroup of $\pi_1(X,x)$ that contains the image of i_1: $\pi_1(U \cap V, x) \to \pi_1(U,x)$.

Corollary 14.9. *If $U \cap V$ is simply connected, then, for any G,*

$$\mathrm{Hom}(\pi_1(X,x),G) \;=\; \mathrm{Hom}(\pi_1(U,x),G) \times \mathrm{Hom}(\pi_1(V,x),G).$$

This means that $\pi_1(X,x)$ is the *free product* of $\pi_1(U,x)$ and $\pi_1(V,x)$.

Exercise 14.10. If $U \cap V$ is simply connected, show that the inclusion mappings j_1 and j_2 are one-to-one.

The following is a useful generalization of Van Kampen's theorem, which can be used to compute the fundamental group of an increasing union of spaces, each of whose fundamental groups in known. The proof is identical to that of the preceding theorem, using the general patching construction of the preceding section.

Suppose a space X is a union of a family of open subspaces X_α, $\alpha \in \mathcal{A}$, with the property that the intersection of any two of these subspaces is in the family. Assume that X and each X_α is path-connected and has a universal covering, and that the intersection of all the X_α contains a point x. When X_β is contained in X_α let $i_{\alpha\beta}$ be the map from $\pi_1(X_\beta,x)$ to $\pi_1(X_\alpha,x)$ determined by the inclusion, and let j_α be the map from $\pi_1(X_\alpha,x)$ to $\pi_1(X,x)$ determined by inclusion.

Theorem 14.11. *With these hypotheses, $\pi_1(X,x)$ is the* direct limit *of the groups $\pi_1(X_\alpha,x)$. That is, for any group G, and any collection of homomorphisms h_α from $\pi_1(X_\alpha,x)$ to G such that $h_\beta = h_\alpha \circ i_{\alpha\beta}$ whenever $X_\beta \subset X_\alpha$, there is a unique homomorphism h from $\pi_1(X,x)$ to G such that $h_\alpha = h \circ j_\alpha$ for all α.* □

The preceding theorem is recovered by taking the family to consist of U, V, and $U \cap V$.

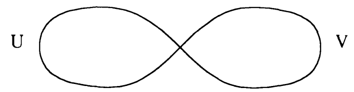
Although this version of Van Kampen's theorem is stated with each subspace X_α open in X, it can often be applied to subspaces that are not open. For example, if each X_α is contained in an open set U_α, of which it is a deformation retract, with $U_\beta \subset U_\alpha$ whenever $X_\beta \subset X_\alpha$, and the hypotheses of Theorem 14.11 apply to these U_α, then $\pi_1(X,x)$ is the direct limit of the groups $\pi_1(X_\alpha,x)$. This follows from the fact that each $\pi_1(X_\alpha,x) \to \pi_1(U_\alpha,x)$ is an isomorphism. Without some such hypotheses, however, the theorem is false. For example, if A and B are copies of a cone over a clamshell (see Exercise 13.19), joined together at the one point where all the circles are tangent, then the spaces A and B are simply connected, and $A \cap B$ is a point, but $A \cup B$ is not simply connected. (In fact, $A \cup B$ is an example of a space that is not simply connected but which has no nontrivial connected coverings.)

Exercise 14.12. Show if a space X is a union of a family of open subspaces X_α such that the intersection of any two sets in the family is also in the family, then H_1X is the direct limit of the groups H_1X_α.

14d. Applications: Graphs and Free Groups

One simple application of the Van Kampen theorem is a result we looked at earlier: the n-sphere S^n is simply connected if $n \geq 2$. To see this now, write the sphere as a union of two hemispheres each homeomorphic to n-dimensional disks, with the intersection homeomorphic to S^{n-1}. It follows from Corollary 14.7 that the fundamental group of S^n is trivial. (The assumption $n \geq 2$ is used to confirm that S^{n-1} is connected.)

Consider next a figure 8:



This is the union X of two circles U and V meeting at a point x. Let γ_1 and γ_2 be loops, one around each circle. The fundamental group of each circle is infinite cyclic, generated by the classes of these loops. It follows that to give a homomorphism from $\pi_1(X,x)$ to a group G is the same as specifying arbitrary elements g_1 and g_2 in G: there is a unique homomorphism from $\pi_1(X,x)$ to G mapping $[\gamma_1]$ to g_1 and $[\gamma_2]$ to g_2. This means that $\pi_1(X,x)$ is the *free group* on the generators $[\gamma_1]$ and $[\gamma_2]$.

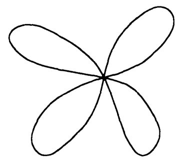
Exercise 14.13. Let $a = [\gamma_1]$ and $b = [\gamma_2]$. Show that every element in $\pi_1(X,x)$ has a unique expression in the form

$$a^{m_0} \cdot b^{m_1} \cdot a^{m_2} \cdot \ . \ . \ . \ . \cdot b^{m_r},$$

where the m_i are integers, all nonzero except perhaps the first and last. The identity element is $e = a^0 b^0$.

The free group on two generators a and b can be constructed directly (and algebraically) as the set of all words of this form, with products defined by juxtaposition of words, canceling to get a legitimate word. It is straightforward (if a little awkward) to show by hand that this forms a group, and to see that it satisfies the above universal property. With the use of the Van Kampen theorem, we can avoid this, by constructing the free group as the fundamental group of this figure 8 space.

The free group F_n on n generators $a_1, \ldots, a_n$ is defined similarly: it is generated by these elements, and, for any group G and any elements $g_1, \ldots, g_n$ in G, there is a unique homomorphism from F_n to G taking a_i to g_i for $1 \leq i \leq n$. Again, it can be constructed purely algebraically using words in these letters, or as a fundamental group:



Exercise 14.14. Verify that the fundamental group of the space obtained by joining n circles at a point is the free group on n generators. Use this to show that the fundamental group of the complement of n points in the plane is free on n generators.

Exercise 14.15. Let X be a connected finite graph. (a) Show that, for any edge between two distinct vertices, X is homotopy equivalent to the graph obtained by removing the edge and identifying its two endpoints. (b) Show that X is homotopy equivalent to the graph obtained by joining n circles at a point, where, if the graph has v vertices and e edges, $n = e - v + 1$. (c) Show that n is the "connectivity" of the graph, i.e., the largest number of edges one can remove from the graph (leaving the vertices), so that what is left remains connected.

One can use this result to give a simple proof of a rather surprising fact about free groups.

Proposition 14.16. *If G is a free group on n generators, and H is a subgroup of G that has finite index d in G, then H is a free group, with $dn - d + 1$ generators.*

Proof. Take G to be the fundamental group of a connected graph X that has v vertices and e edges, with $n = e - v + 1$. For simplicity we assume each edge of X connects two distinct vertices. The subgroup H corresponds to a connected covering $p: Y \to X$ with d sheets, with the fundamental group of Y isomorphic to H. It is not hard to verify that Y is a connected graph. In fact, the d points over each of the vertices of X can be taken as vertices of Y, and (since a covering is trivial over an interval), the d components of the inverse image of each edge are edges of Y. Since Y is a graph, its fundamental group is free, with

$$de - dv + 1 = d(e - v + 1) - d + 1 = dn - d + 1$$

generators. □

If U is the plane domain that is the complement of two points, then U has the figure 8 as a deformation retract. So U has the same fundamental group. In particular, this is not an abelian group. For example, the path $\gamma_1 \cdot \gamma_2 \cdot \gamma_1^{-1} \cdot \gamma_2^{-1}$ is not homotopic to a constant path (although all integrals of all closed 1-forms over this path are trivial).

Problem 14.17. Use the Van Kampen theorem to compute the fundamental groups of the complement of n points (or small disks) in: (1) a two-sphere; (2) a torus; and (3) a projective plane.

Problem 14.18. Describe the fundamental group of $\mathbb{R}^2 \setminus Z$, where Z is the set $\mathbb{Z}$ of all integers, or the set $\mathbb{Z}^2$ of lattice points, or any infinite discrete set.

Problem 14.19. Show that a free group on two generators contains a subgroup that is not finitely generated, in fact, a subgroup that is a free group on an infinite number of variables.

Problem 14.20. Use the Van Kampen theorem to compute the fundamental groups of: (1) the sphere with g handles; (2) the complement of n points in the sphere with g handles; and (3) the sphere with h crosscaps.

PART VIII

COHOMOLOGY AND HOMOLOGY, III

We have seen that coverings of a space can be described by giving coverings on open sets and patching together the coverings over the intersections of these open sets. In particular, one can start with trivial coverings over small open sets, and patch them together, and any covering arises this way. This process is formalized in Chapter 15, and the data that describe such coverings made into Čech cohomology classes. For G-coverings, when G is an abelian group, these classes form a group which we see is "dual" to the homology group, i.e., it is isomorphic to the group of homomorphisms from the first homology group into G.

This is applied to show that the first De Rham group H^1X of an open set in the plane is isomorphic to the dual $\mathrm{Hom}(H_1X, \mathbb{R})$ of the homology group, which gives the culmination of our experience that the homology group and the De Rham group are measuring the same thing about X. This allows us to translate results about homology into corresponding results about cohomology, and in particular to finish the proof of the Mayer–Vietoris theorem for cohomology. (Another proof of this fact will be given in the last chapter of this book.)

Chapter 16, which is optional, contains several miscellaneous variations, applications, and generalities on the same themes, with many of the details left as exercises. The patching construction is used to describe the orientation covering of a manifold. The construction of a covering of a plane domain from a closed 1-form, which follows from the general results of Chapter 15, is carried out here directly using the language of "germs" of functions; in particular, the coverings are seen as graphs of multivalued functions, similar to the polar coordinate covering of Part I. We also describe briefly another cohomology theory.

The last few sections of Chapter 16 generalize the constructions relating coverings with actions of groups and group homomorphisms that were used in Chapters 14 and 15. The added generality, while not needed in this book, may help to make the constructions more understandable by putting them in their natural context. In addition, the exercises (with their hints) carry out proofs of general facts about these constructions, special cases of which were used in Part VII.

CHAPTER 15

Cohomology

15a. Patching Coverings and Čech Cohomology

Since a G-covering is locally trivial, it can be constructed by patching together trivial coverings. In this section we specify exactly what data are needed to carry this out. First, we need to know what this pasting data looks like.

Lemma 15.1. *If $Y = X \times G \to X$ is the trivial G-covering, and $h: X \to G$ is any locally constant function, then the mapping*

$$x \times g \mapsto x \times g \cdot h(x)$$

from Y to Y is an isomorphism of G-coverings, and every isomorphism of G-coverings has this form for a unique locally constant function h.

Proof. It is evident that such a map is continuous and compatible with left multiplication by G, so it is an isomorphism of G-coverings. Conversely, if $\varphi: Y \to Y$ is compatible with left multiplication by G, φ must take each point $x \times g$ to $x \times g \cdot h(x)$ for some $h(x)$ in G. For φ to be continuous, the map $x \mapsto h(x)$ must be locally constant. □

If $p: Y \to X$ is any G-covering, one can find a collection $\mathfrak{U} = \{U_\alpha: \alpha \in \mathcal{A}\}$ of open sets whose union is X such that the restriction of the covering to each U_α is a trivial G-covering. Choose isomorphisms of G-coverings

$$\varphi_\alpha: U_\alpha \times G \xrightarrow{\cong} p^{-1}(U_\alpha).$$

On the overlaps $U_\alpha \cap U_\beta$ that are not empty we have "transition" isomorphisms

$$U_\alpha \cap U_\beta \times G \to p^{-1}(U_\alpha \cap U_\beta) \to U_\alpha \cap U_\beta \times G$$

given by the restriction of φ_α followed by the restriction of ${\varphi_\beta}^{-1}$. These transition isomorphisms have the form $x \times g \mapsto x \times g \cdot g_{\alpha\beta}(x)$ for some (unique) locally constant functions $g_{\alpha\beta}\colon U_\alpha \cap U_\beta \to G$, by the preceding lemma. The collection of these $g_{\alpha\beta}$ satisfy three properties:

(i) $g_{\alpha\alpha} = e$ (the identity in G) for all α;
(ii) $g_{\beta\alpha} = (g_{\alpha\beta})^{-1}$ for all α, β; and
(iii) $g_{\alpha\beta} = g_{\alpha\beta} \cdot g_{\beta\gamma}$ on $U_\alpha \cap U_\beta \cap U_\gamma$ for all α, β, γ.

The first two of these are obvious, and the last follows from the equation ${\varphi_\gamma}^{-1} \circ \varphi_\alpha = {\varphi_\gamma}^{-1} \circ \varphi_\beta \circ {\varphi_\beta}^{-1} \circ \varphi_\alpha$. A collection $\{g_{\alpha\beta}\}$ of locally constant functions[8] satisfying (i)–(iii) is called a *Čech cocycle on $\mathcal{U}$ with coefficients in G*.

Conversely, given $\mathcal{U}$ and a Čech cocycle $\{g_{\alpha\beta}\}$ on $\mathcal{U}$ with coefficients in G, one can use it as "gluing data" to construct a G-covering, together with a trivialization over each U_α, so that the resulting cocycle is the given one. To do this, take the disjoint union of all the products $U_\alpha \times G$ (where G has the discrete topology, i.e., its points are open), and define an equivalence relation by defining $x \times g$ in $U_\alpha \times G$ to be equivalent to $x \times g \cdot g_{\alpha\beta}(x)$ in $U_\beta \times G$ for x in $U_\alpha \cap U_\beta$. Properties (i)–(iii) guarantee precisely that this is an equivalence relation. Define Y to be the set of equivalence classes, with the quotient topology. Since the equivalence relation is compatible with left multiplication by G and with the projections to X, the space Y gets an action of G and a projection p from Y to X so that the resulting maps $U_\alpha \times G \to Y$ are G-maps and compatible with the maps to X. In other words, this is the patching described in §14b, using the transition functions $\vartheta_{\beta\alpha}$, where $\vartheta_{\beta\alpha}(x \times g) = x \times g \cdot g_{\alpha\beta}(x)$.

The cocycle is not uniquely determined by the G-covering, even with a fixed choice of $\mathcal{U}$, since it depends on the choice of trivializations φ_α. But if $\{\varphi_\alpha'\}$ is another choice of trivializations, using the lemma again, there is for each α a unique locally constant function

[8] Note that if the collection U is chosen so that every $U_\alpha \cap U_\beta$ is connected, then these $g_{\alpha\beta}$ are constant, i.e., they are just elements of G. This case will suffice for our applications, and the reader is invited to make this simplifying assumption from the start.

$h_\alpha\colon U_\alpha \to G$ so that the diagram

$$\begin{tikzcd}[column sep=large] & U_\alpha \times G \arrow[dl, "\varphi_\alpha"'] \arrow[dd, "x \times g \,\mapsto\, x \times g \cdot h_\alpha(x)"] \\ \pi^{-1}(U_\alpha) & \\ & U_\alpha \times G \arrow[ul, "\varphi_\alpha'"] \end{tikzcd}$$

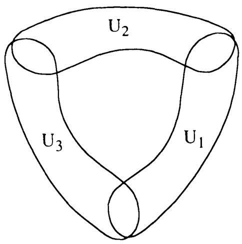
commutes. From this it follows that the cocycle $\{g_{\alpha\beta}'\}$ for the trivializations φ_α' is related to that $\{g_{\alpha\beta}\}$ for the φ_α by the equations

$$g_{\alpha\beta}' = (h_\alpha)^{-1} \cdot g_{\alpha\beta} \cdot h_\beta .$$

Two Čech cocycles $\{g_{\alpha\beta}\}$ and $\{g_{\alpha\beta}'\}$ are said to be *cohomologous* if there are locally constant functions $h_\alpha\colon U_\alpha \to G$ such that $g_{\alpha\beta}' = (h_\alpha)^{-1} \cdot g_{\alpha\beta} \cdot h_\beta$ on $U_\alpha \cap U_\beta$ for all α, β such that $U_\alpha \cap U_\beta$ is nonempty. This is easily checked to be an equivalence relation, and the equivalence classes are called *Čech cohomology classes* on $\mathcal{U}$ *with coefficients in* G. The set of these is denoted $H^1(\mathcal{U}; G)$.

Exercise 15.2. Verify that the G-coverings constructed from cohomologous cocycles are isomorphic.

Exercise 15.3. Let $\mathcal{U}$ be a covering of an annulus by three sets homeomorphic to disks, as shown:



For g in G let $c(g)$ be the cocycle determined by setting $g_{12} = g_{23} = e$ and $g_{31} = g$. (a) Show that every Čech cocycle on $\mathcal{U}$ with coefficients in G is cohomologous to such a cocycle, and show that $c(g)$ is cohomologous to $c(g')$ if and only if g and g' are conjugate in G. This gives a bijection between the set of conjugacy classes in G and $H^1(\mathcal{U}; G)$. (b) If G is cyclic of prime order, show that the coverings corresponding to any two distinct elements that are not the identity in G are isomorphic as coverings, but not as G-coverings.

What we have done in this section amounts to setting up a one-to-one correspondence between $H^1(\mathcal{U};G)$ and

$$\{G\text{-coverings of } X \text{ that are trivial over each } U_\alpha\}/\cong.$$

If the U_α are chosen to be simple open sets like disks or rectangles, or any simply connected and locally path-connected open sets, then every covering is trivial over them (by Corollary 13.8), so we will have all G-coverings of X classified by Čech cohomology classes. Thus we have:

Proposition 15.4. *If each U_α is simply connected and locally path-connected, then the set of G-coverings of X, up to isomorphism, is in one-to-one correspondence with the Čech cohomology set $H^1(\mathcal{U};G)$.*

Now suppose X is a connected space that has a universal covering. From Exercise 14.3 we know that the set of G-coverings up to isomorphism is in one-to-one correspondence with the set of homomorphisms from $\pi_1(X,x)$ to G, up to conjugacy. Putting this together, and continuing to assume that the open sets in $\mathcal{U}$ are simply connected, we have

Corollary 15.5. *With these assumptions, there is a bijection*

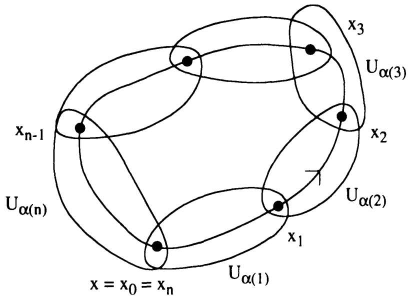
$$H^1(\mathcal{U};G) \leftrightarrow \mathrm{Hom}(\pi_1(X,x),G)/\mathit{conjugacy}.$$

15b. Čech Cohomology and Homology

If we want to have a well-defined homomorphism from $\pi_1(X,x)$ to G, not just up to conjugacy, we know that we must consider coverings $Y \to X$ together with a base point y in Y over x. To do this with the patching, for those open sets U_α that contain x, fix the local trivializations $\varphi_\alpha\colon U_\alpha \times G \to p^{-1}(U_\alpha)$ so that $\varphi_\alpha(x \times e) = y$. The corresponding cocycle then has $g_{\alpha\beta}(x) = e$ whenever x is in $U_\alpha \cap U_\beta$; another cocycle defines an isomorphic covering preserving base points when it is of the form $(h_\alpha)^{-1} \cdot g_{\alpha\beta} \cdot h_\beta$, with $h_\alpha\colon U_\alpha \to G$ locally constant with the restriction that $h_\alpha(x) = e$ whenever x is in U_α. Such equivalence classes form a set we can denote by $H^1(\mathcal{U},x;G)$. This sets up bijections

$$\begin{aligned} H^1(\mathcal{U},x;G) &\leftrightarrow \{G\text{-coverings with base point}\}/\cong \\ &\leftrightarrow \mathrm{Hom}(\pi_1(X,x),G). \end{aligned}$$

We will need a prescription to make the homomorphism ρ from $\pi_1(X,x)$ to G explicit, when the covering is given as above by a cocycle. As one should expect, the answer is the product of values of the transition functions as one moves around a path, just as the winding number is the sum of changes in angle. If γ is a loop at x, subdivide the unit interval, $0=t_0<t_1< \ldots <t_n=1$ such that γ maps each subinterval $[t_{i-1},t_i]$ into one of the open sets $U_{\alpha(i)}$. Let $x_i=\gamma(t_i)$.



Lemma 15.6. *The homomorphism $\rho\colon \pi_1(X,x)\to G$ corresponding to the G-covering $p\colon Y\to X$ takes the homotopy class of γ to*

$$\rho([\gamma]) = h_1\cdot h_2\cdot \ldots \cdot h_{n-1}\cdot h_n,$$

where $h_i=g_{\alpha(i)\alpha(i+1)}(x_i)$ for $1\le i\le n-1$, and $h_n=g_{\alpha(n)\alpha(1)}(x_n)$.

Proof. Let $\tilde\gamma$ be the lift of γ starting at y. We must compute $\tilde\gamma(t_i)$ for each i. At the start, $\tilde\gamma(0)=y=\varphi_{\alpha(1)}(x\times e)$. For $0\le t\le t_1$, by continuity, $\tilde\gamma(t)=\varphi_{\alpha(1)}(\gamma(t)\times e)$. So

$$\tilde\gamma(t_1) = \varphi_{\alpha(1)}(x_1\times e) = \varphi_{\alpha(2)}(x_1\times e\cdot g_{\alpha(1)\alpha(2)}(x_1)) = \varphi_{\alpha(2)}(x_1\times h_1).$$

Going along each piece of the path in the same way, we find

$$\tilde\gamma(t_i) = \varphi_{\alpha(i)}(x_i\times h_1\cdot \ldots \cdot h_{i-1}) = \varphi_{\alpha(i+1)}(x_i\times h_1\cdot \ldots \cdot h_i).$$

At the end, this gives

$$\begin{aligned}\tilde\gamma(1) &= \varphi_{\alpha(1)}(x\times h_1\cdot \ldots \cdot h_n) = (h_1\cdot \ldots \cdot h_n)\cdot \varphi_{\alpha(1)}(x\times e)\\ &= (h_1\cdot \ldots \cdot h_n)\cdot y.\end{aligned}$$

Since $\tilde\gamma(1)=\rho([\gamma])\cdot y$ by definition, the lemma follows. □

Now we specialize to the case where G is an abelian group. In this case any homomorphism from $\pi_1(X,x)$ to G must send any commutator $aba^{-1}b^{-1}$ in $\pi_1(X,x)$ to the identity of G, so it must define a homomorphism on the abelian quotient group $\pi_1(X,x)_{\text{abel}}$. By Proposition 12.22, we have an isomorphism $\pi_1(X,x)_{\text{abel}} \cong H_1X$. And conjugate homomorphisms to an abelian group must be equal. Putting all this together, and assuming X and $\mathcal{U}$ are as in Proposition 15.4, we have:

Corollary 15.7 (Hurewicz). *If G is abelian, there are canonical bijections*

$$\begin{aligned} H^1(\mathcal{U};G) &\leftrightarrow \operatorname{Hom}(\pi_1(X,x),G) \leftrightarrow \{G\text{-coverings of } X\}/\cong \\ &\leftrightarrow \operatorname{Hom}(\pi_1(X,x)/[\pi_1(X,x),\pi_1(X,x)],G) \\ &\leftrightarrow \operatorname{Hom}(H_1X,G). \end{aligned}$$

For a space X that may not be connected, we have:

Corollary 15.8. *If X is a locally simply connected space, and $\mathcal{U}$ is a collection of simply connected open sets whose union is X, then there are canonical bijections*

$$H^1(\mathcal{U};G) \leftrightarrow \{G\text{-coverings of } X\}/\cong \;\leftrightarrow \operatorname{Hom}(H_1X,G).$$

Proof. To give a G-covering of X is the same as giving a G-covering of each connected component of X, and to give a homomorphism from H_1X to G is the same as giving a homomorphism from the first homology group of each connected component of X to G. So the result follows from Corollary 15.7. □

If G is abelian, the Čech cocycles on $\mathcal{U}$ with coefficients in G form an abelian group, by multiplication: $\{g_{\alpha\beta}\}\cdot\{g_{\alpha\beta}'\} = \{g_{\alpha\beta}\cdot g_{\alpha\beta}'\}$. Call a cocycle $\{g_{\alpha\beta}\}$ a *coboundary* if there are $h_\alpha\colon U_\alpha \to G$ for each α such that $g_{\alpha\beta} = h_\alpha^{-1}\cdot h_\beta$.

Exercise 15.9. Assume that G is abelian. Show that the coboundaries form a subgroup of the cocycles, and that $H^1(\mathcal{U};G)$ is the quotient group of cocycles modulo coboundaries. (Caution: If G is not abelian, $H^1(\mathcal{U};G)$ has no natural group structure, see Exercise 15.3.) If G is abelian, give $\operatorname{Hom}(H,G)$ the structure of an abelian group, for any group H. Show that the bijections of Corollary 15.7 are isomorphisms of abelian groups.

Exercise 15.10. Define the 0th Čech group $H^0(\mathcal{U};G)$ for an open covering $\mathcal{U}$ of a space X with coefficients in a group by defining a class (or cocycle—there is no equivalence relation) to be a collection of locally constant functions $g_\alpha\colon U_\alpha \to G$ such that $g_\alpha = g_\beta$ on $U_\alpha \cap U_\beta$.

Show that $H^0(\mathcal{U}; G)$ is the direct product of copies of G, one for each connected component of X. In particular, for X open in the plane, $H^0(\mathcal{U}; \mathbb{R})$ is isomorphic to the space $H^0(X)$ of locally constant functions.

15c. De Rham Cohomology and Homology

We want to compare the De Rham group H^1X with the first homology group H_1X, when X is an open set in the plane. If ω is a closed 1-form, and γ a closed 1-chain, we defined the integral $\int_\gamma \omega$ in Chapter 9. For fixed ω, this map $\gamma \mapsto \int_\gamma \omega$ is a homomorphism from the group Z_1X of closed 1-chains to $\mathbb{R}$. Proposition 9.11 says that this integral is the same for homologous 1-chains. In other words, the map vanishes on 1-boundaries B_1X, and therefore defines a homomorphism from H_1X to $\mathbb{R}$. Let us denote this homomorphism, at least temporarily, by φ_ω, so that $\varphi_\omega([\gamma]) = \int_\gamma \omega$.

The set $\mathrm{Hom}(H_1X, \mathbb{R})$ of homomorphisms from H_1X to $\mathbb{R}$ has a natural structure of vector space: the sum $\varphi + \psi$ of two homomorphisms is defined by $(\varphi + \psi)([\gamma]) = \varphi([\gamma]) + \psi([\gamma])$, and multiplication of φ by a scalar r by $(r \cdot \varphi)([\gamma]) = r \cdot \varphi([\gamma])$. (Note that this works for any group in place of H_1X.) The above map that assigns φ_ω to the closed 1-form ω is a linear mapping of vector spaces

$$\{\text{closed 1-forms on } X\} \to \mathrm{Hom}(H_1X, \mathbb{R}).$$

This follows from the equation $\int_\gamma (r_1\omega_1 + r_2\omega_2) = \int_\gamma r_1\omega_1 + \int_\gamma r_2\omega_2$. If the 1-form ω is exact, the homomorphism φ_ω is zero. In fact, if $\omega = df$, and γ is any 1-chain, with boundary $\partial\gamma = \Sigma m_i P_i$, then

$$\int_\gamma df = \sum m_i f(P_i).$$

It follows that the above mapping vanishes on the subspace of exact 1-forms, so it defines a linear map on the quotient space H^1X. That is, we have a natural linear map of vector spaces

$$H^1X \to \mathrm{Hom}(H_1X, \mathbb{R}), \qquad [\omega] \mapsto \left([\gamma] \mapsto \int_\gamma \omega\right).$$

If $\alpha \in H^1X$ and $a \in H_1X$, we may write simply $\int_a \alpha$ in place of $\int_\gamma \omega$, where ω is a representative of α and γ a representative of a. The goal of this section is to show that this mapping is always an isomorphism.

Theorem 15.11. *If X is an open set in the plane, then the canonical homomorphism*

$$H^1X \to \operatorname{Hom}(H_1X, \mathbb{R})$$

is an isomorphism.

The fact that this map is one-to-one is not hard to see, for if a closed 1-form ω has all integrals $\int_\gamma \omega$ vanishing for all closed paths γ on X, we know from Proposition 1.8 that ω is exact. The fact that every homomorphism comes from a 1-form, however, will take some work. As a warm-up you may consider the special case:

Exercise 15.12. Show that the homomorphism $H^1X \to \operatorname{Hom}(H_1X, \mathbb{R})$ is an isomorphism when X is multiply connected.

Exercise 15.13. Show that, if U is an open subset of X, the diagram

$$\begin{tikzcd} H^1X \arrow[r] \arrow[d] & H^1U \arrow[d] \\ \operatorname{Hom}(H_1X, \mathbb{R}) \arrow[r] & \operatorname{Hom}(H_1U, \mathbb{R}) \end{tikzcd}$$

commutes, where the map on the bottom is determined by the map from H_1U to H_1X.

In order to prove the theorem, we specialize the results of §15b to the case where $G = \mathbb{R}$ is the additive group of real numbers, and X is an open set in the plane. Let $\mathcal{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$ be a collection of open rectangles whose union is X. We may find such a collection so that any point in X is contained in only finitely many U_α (see Lemma A.20 and Lemma 24.10). By Corollary 15.8 we have a bijection between $\operatorname{Hom}(H_1X, \mathbb{R})$ and $H^1(\mathcal{U}; \mathbb{R})$. To prove the theorem we will construct a map from $H^1(\mathcal{U}; \mathbb{R})$ to the De Rham group H^1X, and then show that all these maps are compatible and isomorphisms.

An element of $H^1(\mathcal{U}; \mathbb{R})$ is determined by a Čech cocycle $\{g_{\alpha\beta}\}$, where the $g_{\alpha\beta}$ are locally constant functions on $U_\alpha \cap U_\beta$. We want to produce from this a closed 1-form ω on X, well defined up to the addition of an exact 1-form. If we can find some $\mathscr{C}^\infty$ functions f_α on U_α so that $f_\alpha - f_\beta = g_{\alpha\beta}$ on $U_\alpha \cap U_\beta$, then $df_\alpha = df_\beta$ on the overlaps, so there is a unique 1-form ω on U that is df_α on each U_α. This will be the 1-form we are after. The existence of such functions f_α follows from a general lemma:

Lemma 15.14. *Let $\{f_{\alpha\beta}\}$ be a collection of $\mathscr{C}^\infty$ functions, $f_{\alpha\beta}$ on $U_\alpha \cap U_\beta$,* *satisfying the cocycle conditions*: (*i*) $f_{\alpha\alpha} = 0$; (*ii*) $f_{\beta\alpha} = -f_{\alpha\beta}$; *and* (*iii*) $f_{\alpha\gamma} = f_{\alpha\beta} + f_{\beta\gamma}$ *on* $U_\alpha \cap U_\beta \cap U_\gamma$. *Then there are* $\mathscr{C}^\infty$ *functions* f_α *on* U_α, *for all* α, *such that*

$$f_\alpha - f_\beta = f_{\alpha\beta} \qquad on \quad U_\alpha \cap U_\beta .$$

Proof. In order to solve these equations $f_\alpha - f_\beta = f_{\alpha\beta}$, we use the existence of a partition of unity subordinate to the covering $\mathcal{U}$ (Appendix B2). This says there are $\mathscr{C}^\infty$ functions φ_α on X, with the closure of the support of φ_α (in X) contained in U_α, with only finitely many φ_α nonzero in a neighborhood of any point, and with $\Sigma_\alpha \varphi_\alpha \equiv 1$. For each α and β define a $\mathscr{C}^\infty$ function $h_{\alpha\beta}$ on U_α by the formula

$$h_{\alpha\beta} = \begin{cases} \varphi_\beta \cdot f_{\alpha\beta} & \text{on } U_\alpha \cap U_\beta, \\ 0 & \text{on } U_\alpha \setminus U_\alpha \cap U_\beta . \end{cases}$$

Now set $f_\alpha = \Sigma_\beta h_{\alpha\beta}$. It is an easy exercise to verify that $h_{\alpha\beta}$ and f_α are $\mathscr{C}^\infty$ functions. To complete the proof, we calculate:

$$\begin{aligned} f_\alpha - f_\beta &= \sum_\gamma (\varphi_\gamma f_{\alpha\gamma} - \varphi_\gamma f_{\beta\gamma}) = \sum_\gamma (\varphi_\gamma (f_{\alpha\gamma} - f_{\beta\gamma}) \\ &= \sum_\gamma \varphi_\gamma (f_{\alpha\beta}) = \Big(\sum_\gamma \varphi_\gamma\Big) f_{\alpha\beta} = 1 \cdot f_{\alpha\beta} = f_{\alpha\beta} . \end{aligned}$$ □

We must verify that the construction made before the lemma gives a well-defined map from $H^1(\mathcal{U}; \mathbb{R})$ to H^1X. Suppose first that $\{f_\alpha'\}$ is another collection of functions, with f_α' on U_α such that $f_\alpha' - f_\beta' = g_{\alpha\beta}$ on $U_\alpha \cap U_\beta$, giving rise to the 1-form ω' that is df_α' on U_α. The functions $f_\alpha' - f_\alpha$ on U_α agree on the overlaps, so define a function f on X; and $\omega' - \omega = df$, so ω' and ω define the same element of H^1X. We must also show that if $\{g_{\alpha\beta}'\}$ is a Čech cocycle that is cohomologous to $\{g_{\alpha\beta}\}$, then they determine the same class in H^1X. We are given locally constant functions h_α on U_α such that

$$g_{\alpha\beta}' = -h_\alpha + g_{\alpha\beta} + h_\beta \qquad \text{on} \quad U_\alpha \cap U_\beta .$$

If $f_\alpha - f_\beta = g_{\alpha\beta}$ on $U_\alpha \cap U_\beta$, then $(f_\alpha - h_\alpha) - (f_\beta - h_\beta) = g_{\alpha\beta}'$ on $U_\alpha \cap U_\beta$, and since each h_α is locally constant, $d(f_\alpha - h_\alpha) = df_\alpha$, so they define the same 1-form.

Summarizing what we have so far, we have maps

$$\begin{tikzcd} H^1X \arrow[r] & \mathrm{Hom}(H_1X, \mathbb{R}) \arrow[d, leftrightarrow] \\ H^1(\mathcal{U}; \mathbb{R}) \arrow[u] \arrow[r, leftrightarrow] & \{\mathbb{R}\text{-coverings}\}/\cong . \end{tikzcd}$$

We want to show that all of these maps are bijections, and that going once clockwise around the diagram, starting at any place, is the identity map on that vector space. We have proved that the top horizontal map in this diagram is one-to-one. It therefore suffices to show that a trip around the diagram, starting with an element ρ in $\mathrm{Hom}(H_1X, \mathbb{R})$, takes ρ to itself. Let $\{f_{\alpha\beta}\}$ be a cocycle for the covering $p_\rho\colon Y_\rho \to X$ defined from this homomorphism. Use Lemma 15.14 to write $f_{\alpha\beta} = f_\alpha - f_\beta$, and let ω be the 1-form that is df_α on U_α. It follows from Proposition 12.22 that H_1X is generated by the classes of closed paths in X. To conclude the proof, it therefore suffices to show that

$$\int_\gamma \omega = \rho([\gamma])$$

when γ is any closed path in X. Both sides depend only on the connected component of X containing the image of γ, so we may assume X is connected. Subdivide the path as in Lemma 15.6, and let γ_i denote the restriction of γ to the ith piece $[t_{i-1}, t_i]$. By Lemma 15.6, $\rho([\gamma]) = \sum_{i=1}^n h_i$, where $h_i = f_{\alpha(i)\alpha(i+1)}(x_i)$ for $1 \le i \le n$, with $\alpha(n+1) = \alpha(1)$. Therefore,

$$\begin{aligned}\rho([\gamma]) &= \sum_{i=1}^n f_{\alpha(i)}(x_i) - f_{\alpha(i+1)}(x_i) = \sum_{i=1}^n f_{\alpha(i)}(x_i) - f_{\alpha(i)}(x_{i-1}) \\ &= \sum_{i=1}^n \int_{\gamma_i} df_{\alpha(i)} = \sum_{i=1}^n \int_{\gamma_i} \omega = \int_\gamma \omega .\end{aligned}$$

This completes the proof of Theorem 15.11. □

Corollary 15.15. *There are canonical bijections*

$$H^1X \leftrightarrow \mathrm{Hom}(H_1X, \mathbb{R}) \leftrightarrow \{\mathbb{R}\text{-}coverings\}/\cong \leftrightarrow H^1(\mathcal{U}; \mathbb{R}).$$

Exercise 15.16. Verify directly that the maps of this corollary are linear maps between vector spaces.

The theorem can be used to calculate the De Rham group of more complicated sets in the plane than multiply connected sets, even when it may be difficult to write down differential forms explicitly:

Problem 15.17. (a) Show that if X is the complement of the set $\mathbb{N}$ of nonnegative integers, then H^1X is isomorphic to the space of all infinite sequences $(a_0, a_1, \ldots)$ of real numbers. In particular, H^1X is infinite dimensional. (b) Compute H^1X when X is the complement of the set $(0, 1, 1/2, 1/3, 1/4, \ldots)$. (c) What if X is the complement of any infinite discrete set?

15d. Proof of Mayer–Vietoris for De Rham Cohomology

Now that we have explicitly identified the De Rham cohomology group with the dual of the homology group, we can finish the proof of the Mayer–Vietoris theorem stated in §10d for the De Rham groups. What was missing is the assertion MV(vi) that the map denoted "$-$" from $H^1U \oplus H^1V$ to $H^1(U \cap V)$ is surjective. We also have a linear map

$$H^1U \oplus H^1V \to \mathrm{Hom}(H_1(U) \oplus H_1(V), \mathbb{R})$$

that takes a pair (α, β) to the homomorphism that takes a pair (a, b) in $H_1(U) \oplus H_1(V)$ to $\int_a \alpha + \int_b \beta$. It is a simple exercise to verify, using Theorem 15.11, that this map is also an isomorphism.

The homomorphism "$-$" from $H_1(U \cap V)$ to $H_1U \oplus H_1V$ determines a homomorphism

$$\mathrm{Hom}(H_1(U) \oplus H_1(V), \mathbb{R}) \to \mathrm{Hom}(H_1(U \cap V), \mathbb{R}),$$

and it is a general fact, proved in Appendix C (Lemma C.10), that, since the map that determines it is one-to-one, this map is surjective. Consider the diagram

$$\begin{array}{ccc}
H^1U \oplus H^1V & \longrightarrow & H^1(U \cap V) \\
\downarrow & & \downarrow \\
\mathrm{Hom}(H_1(U) \oplus H_1(V), \mathbb{R}) & \longrightarrow & \mathrm{Hom}(H_1(U \cap V), \mathbb{R})
\end{array}$$

where the maps are as just described. We know that each of the vertical maps is an isomorphism, and that the bottom horizontal map is surjective. It follows that the top horizontal map is also surjective, which is the assertion to be proved, provided we check that the diagram commutes. But this is entirely straightforward. Given a class (α, β) in $H^1U \oplus H^1V$, going either way around the diagram, it goes to the homomorphism that takes an element γ in $H_1(U \cap V)$ to the number $\int_\gamma \alpha - \int_\gamma \beta$. □

In fact, the whole Mayer–Vietoris sequence in cohomology is dual to, and can be deduced from, the Mayer–Vietoris sequence in homology. The 0th De Rham group H^0X is dual to the 0th homology group H_0X, as follows. For any function f on X and any 0-cycle $\zeta = \Sigma m_i P_i$ on X, define $f(\zeta) = \Sigma m_i f(P_i)$. If f is locally constant, and ζ is a 0-boundary, then $f(\zeta) = 0$, as follows from the fact that f must take the same values at the two endpoints of any path. This determines a linear map

$$H^0X \to \mathrm{Hom}(H_0X, \mathbb{R}), \qquad f \mapsto ([\zeta] \mapsto f(\zeta)).$$

Exercise 15.18. Use the fact that H_0X is the free abelian group on the connected components of X to prove that this map from H^0X to $\mathrm{Hom}(H_0X, \mathbb{R})$ is an isomorphism. Show that the reduced groups are also dual: $\widetilde{H}^0X \cong \mathrm{Hom}(\widetilde{H}_0X, \mathbb{R})$.

For simplicity now, for any abelian group A, write A^* for $\mathrm{Hom}(A, \mathbb{R})$. Note that a homomorphism $A \to B$ determines a linear map $B^* \to A^*$. A direct sum decomposition $A = B \oplus C$ determines a direct sum decomposition $A^* = B^* \oplus C^*$. The dual to the Mayer–Vietoris homology sequence is the sequence of vector spaces:

$$0 \longrightarrow H_0(U \cup V)^* \xrightarrow{+} H_0U^* \oplus H_0V^* \xrightarrow{-} H_0(U \cap V)^*$$

$$\xrightarrow{\partial^*} H_1(U \cup V)^* \xrightarrow{+} H_1U^* \oplus H_1V^* \xrightarrow{-} H_1(U \cap V)^* \longrightarrow 0$$

The fact that the homology sequence is exact (Theorem 10.5) implies that this is an exact sequence of vector spaces (see Appendix C2). The isomorphisms from the preceding section can be used to map each term in the Mayer–Vietoris cohomology sequence to the corresponding term in the above sequence:

$$\begin{array}{ccccccccc}
0 \longrightarrow & H^0(U \cup V) & \xrightarrow{+} & H^0U \oplus H^0V & \xrightarrow{-} & H^0(U \cap V) & \xrightarrow{\delta} & H^1(U \cup V) \\
& \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 \longrightarrow & H_0(U \cup V)^* & \xrightarrow{+} & H_0U^* \oplus H_0V^* & \xrightarrow{-} & H_0(U \cap V)^* & \xrightarrow{\partial^*} & H^1(U \cup V)^*
\end{array}$$

$$\begin{array}{ccccc}
\xrightarrow{+} & H^1U \oplus H^1V & \xrightarrow{-} & H^1(U \cap V) & \longrightarrow 0 \\
& \downarrow & & \downarrow & \\
\xrightarrow{+} & H_1U^* \oplus H_1V^* & \xrightarrow{-} & H_1(U \cap V)^* & \longrightarrow 0.
\end{array}$$

Problem 15.19. (a) Show that each square in the above diagram commutes. (b) Use this to deduce the exactness of the cohomology sequence from the exactness of the homology sequence.

CHAPTER 16

Variations

16a. The Orientation Covering

Every manifold M has a canonical two-sheeted covering $p: \tilde{M} \to M$, called the *orientation covering*, whose fiber over a point P is the two ways to orient M at P. Čech cocycles provide a convenient way to construct this covering. Let $G = \{\pm 1\}$ be the group of order two, and take an open covering $\mathcal{U} = \{U_\alpha\}$ of M to be images of the coordinate charts $\varphi_\alpha: V_\alpha \to U_\alpha \subset M$, with V_α open in $\mathbb{R}^n$. The Jacobian determinant of the change of coordinates from V_α to V_β has a locally constant sign, which gives a locally constant function from $U_\alpha \cap U_\beta$ to $\{\pm 1\}$. The chain rule for Jacobians implies that this is a cocycle. Define $p: \tilde{M} \to M$ to be the resulting $\{\pm 1\}$-covering. An *orientation* of M can be defined to be a section of this covering, i.e., a continuous mapping $\sigma: M \to \tilde{M}$ such that $p \circ \sigma$ is the identity map on M.

Exercise 16.1. (a) Show that this orientation covering is independent of the choice of coordinate charts. (b) If M is connected, show that $\tilde{M}$ has one or two connected components: one if M is nonorientable, two if M is orientable. (c) Show that $\tilde{M}$ is always orientable.

Exercise 16.2. When $M = \mathbb{RP}^2$ show that the orientation covering is isomorphic to the covering $S^2 \to \mathbb{RP}^2$. Do the same for $\mathbb{RP}^n$, $n > 2$. Identify the orientation covering for the Klein bottle.

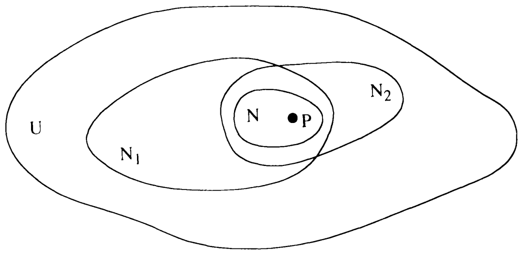
16b. Coverings from 1-Forms

From Corollary 15.7 and Theorem 15.11 we have a canonical bijection

$$H^1X \leftrightarrow \{\mathbb{R}\text{-coverings of } X\}/\cong.$$

It follows that there is an $\mathbb{R}$-covering $p_\omega: X_\omega \to X$ of an open set X in the plane corresponding to a closed 1-form ω on X, with two 1-forms giving isomorphic coverings when their difference is exact. To construct this covering, following the procedure given in §15c, one takes a collection $\mathcal{U}$ of open sets U_α on each of which $\omega = df_\alpha$ for some function f_α, and then uses the transition functions $g_{\alpha\beta} = f_\alpha - f_\beta$ to patch together trivial $\mathbb{R}$-coverings of each U_α into a global $\mathbb{R}$-covering of X. In this section we describe a more direct way to construct this covering.

The idea is to put all graphs of functions f on open subsets of X such that $df = \omega$ together into one big covering space. A good language for describing this is that of germs. A *germ* of a $\mathcal{C}^\infty$ function *at* a point P in X is an equivalence class of $\mathcal{C}^\infty$ functions defined in neighborhoods of P. Two functions f_1 on a neighborhood N_1 and f_2 on N_2 are defined to be equivalent if there is a neighborhood N of P contained in $N_1 \cap N_2$ such that f_1 and f_2 are equal on N.



These germs can be added, multiplied, and differentiated, just like functions on definite open sets.

Exercise 16.3. Verify that this is an equivalence relation, and that these operations preserve equivalence classes.

A germ at P has a value at P (the value $f(P)$ of any representative f), but it does not have a value at any other point. The idea is simply that we only care about functions near the point P, and we allow ourselves to shrink the neighborhoods arbitrarily. Given a function f

on a neighborhood of P, the equivalence class containing it is called the germ *defined by f at P*.

Since derivatives of germs make sense, it makes sense to say that the *differential* of a germ at a point is a given 1-form; if the 1-form is $\omega = p\,dx + q\,dy$, the differential of a germ of f at P is ω if

$$\frac{\partial f}{\partial x}(P) = p(P) \quad \text{and} \quad \frac{\partial f}{\partial y}(P) = q(P).$$

If ω is a closed 1-form on X, define X_ω to be the set of all germs of functions at points of X whose differential is ω. We will make this into a topological space, so that the map $p_\omega \colon X_\omega \to X$, that takes a germ to the point at which it is a germ, is a covering map. For each open set N in X and function f on N such that $df = \omega$ on N, define a basic open set N_f in X_ω:

$$N_f \;=\; \{\text{germs at points of } N \text{ defined by } f\}.$$

A set in X_ω is defined to be open if it is a union of basic sets N_f.

If N is an open set in X such that ω is exact on N (for example, any open disk or rectangle in X), then $p_\omega{}^{-1}(N)$ is a disjoint union of the open sets N_f, where f runs through all functions on N whose differential is ω. The point is that if f is one such function, then all others, on N or on any open subset of N, are of the form $f + c$ for some locally constant function c. This shows that the covering is trivial over N, and, since any point in X has such neighborhoods, this shows that we have a covering space:

$$p_\omega \colon X_\omega \to X.$$

In addition, this is an $\mathbb{R}$-covering. The action of a real number on X_ω takes a germ to the sum of the germ and the real number. The preceding discussion shows that this is an even action of $\mathbb{R}$ on X_ω, and that the covering is an $\mathbb{R}$-covering.

Exercise 16.4. Verify that when $X = \mathbb{R}^2 \setminus \{(0,0)\}$ and $\omega = d\vartheta$, then the polar coordinate covering of X is a connected component of X_ω.

Choosing any function f on any connected open subset N of X with $df = \omega$ on N determines a connected component of X_ω. Namely, this is the connected component containing the open set N_f.

Exercise 16.5. Show that this component is the union of all open sets of the form $N'_{f'}$ with N' a connected open set in X and $df' = \omega$ on N' such that there is a chain of connected open sets $N = N_0$,

$N_1, \ldots, N_r = N'$ in X, with functions f_i on N_i such that $df_i = \omega$, and $f_i = f_{i+1}$ on $N_i \cap N_{i+1}$.

In general, there is a natural bijection between X_ω and the product $X \times \mathbb{R}$, that takes a germ at P to the pair $P \times r$, where r is the value of the germ at P. If $\mathbb{R}$ is has its usual topology, this bijection is continuous, but it is *not* a homeomorphism. In the polar coordinate example, each connected component of X_ω maps to a closed submanifold of the product $X \times \mathbb{R}$; these are the graphs of the multivalued functions "$\vartheta + c$." In general, however, this need not be the case. Problem 6.22 shows that, when X is the complement of two points P and Q, and ω is a linear combination $r\,d\vartheta_P + s\,d\vartheta_P$ with r/s irrational, the sheets of a component of X_ω can come arbitrarily close together, if regarded as a subset of $X \times \mathbb{R}$. With this understanding, the connected components of X_ω can be regarded as the multivalued functions on X whose differential is ω—but their graphs need not be closed in $X \times \mathbb{R}$.

Problem 16.6. (a) Show that this covering $p_\omega\colon X_\omega \to X$ is isomorphic to the covering constructed with cocycles. (b) Show directly that 1-forms that differ by an exact form determine isomorphic coverings.

Problem 16.7. Assume X is connected, and let Y be a connected component of X_ω. Show that $Y \to X$ is a G-covering, where G is the period module, i.e., G is the subgroup of $\mathbb{R}$ generated by the periods of ω; these periods are the numbers $\int_\gamma \omega$, as γ varies over loops at x (or equivalently, closed 1-chains on X).

16c. Another Cohomology Group

In addition to the De Rham and Čech cohomology groups, there are cohomology groups $H^0(X; G)$ and $H^1(X; G)$ that one can assign to any topological space X, where G can be any abelian group. Lacking smooth functions and 1-forms to evaluate on points and paths, one makes these groups directly out of objects that are defined by their values on points and paths. Define a *0-cochain* to be an arbitrary function from X to G. Define a *1-cochain* to be a function from the set of all continuous paths on X to G, with the requirement that the function takes all constant paths to the identity element 0 in G. Note that, unlike the case of 0-chains or 1-chains, cochains can assign nonzero elements of G to arbitrarily many elements. Since the set of functions from any set to an abelian group G is an abelian group (by adding the values in G), the 0-cochains form a group denoted $C^0(X;G)$, and the 1-cochains form a group $C^1(X;G)$.

If $a\colon X \to G$ is a 0-cochain, define the *coboundary* δa to be the 1-cochain whose value on a path γ is $a(\gamma(1)) - a(\gamma(0))$. Define the 0th *cohomology group* $H^0(X;G)$ be the set of 0-cochains whose boundary is zero.

Exercise 16.8. Show that giving an element of $H^0(X;G)$ is equivalent to giving a function from the set of path-connected components of X to G. Deduce an isomorphism $H^0(X;G) \cong \mathrm{Hom}(H_0X, G)$.

If c is a 1-cochain, it defines a homomorphism from the 1-chains C_1X to G by the formula $c(\Sigma n_i\gamma_i) = \Sigma n_i c(\gamma_i)$. A 1-cochain c is called a 1-*cocycle* if $c(\partial\Gamma) = 0$ for all continuous Γ: $[0,1] \times [0,1] \to X$. The 1-cocycles form a subgroup $Z^1(X;G)$ of $C^1(X;G)$. Every coboundary δa is a 1-cocycle, so the 1-coboundaries form a subgroup $B^1(X;G)$ of $Z^1(X;G)$. The quotient group is the *first cohomology group*

$$H^1(X;G) \;=\; Z^1(X;G)/B^1(X;G).$$

There is a natural homomorphism from $H^1(X;G)$ to $\mathrm{Hom}(X_1X, G)$, that takes the class of a 1-cocycle c to the homomorphism that takes the homology class of a 1-cycle γ to the element $c(\gamma)$.

Exercise 16.9. Verify that this is a well-defined homomorphism.

Proposition 16.10. *The homomorphism $H^1(X;G) \to Hom(H_1X, G)$ is an isomorphism.*

Proof. To give a homomorphism from $H_1X = Z_1X/B_1X$ to G is equivalent to giving a homomorphism from Z_1X to G that vanishes on B_1X; that is,

$$\mathrm{Hom}(H_1X, G) \;=\; \mathrm{Kernel}(\mathrm{Hom}(Z_1X, G) \to \mathrm{Hom}(B_1X, G)).$$

We have homomorphisms

$$B^1(X;G) \;\to\; Z^1(X;G) \;\to\; \mathrm{Hom}(Z_1X, G) \;\to\; \mathrm{Hom}(B_1X, G),$$

where the middle map takes the cocycle c to the homomorphism $\gamma \to c(\gamma)$. The assertion of the proposition is easily seen to be equivalent to the assertion that this sequence is exact at the two middle groups.

The fact that the image of each map is contained in the next is the content of the preceding exercise. To prove the opposite inclusions, we need to choose arbitrarily a point x_α in each path component X_α of X, and choose for each point y in X an arbitrary path τ_y starting at the chosen point x_α of the component containing y and ending at y.

Suppose c in $Z^1(X;G)$ maps to the zero homomorphism from Z_1X to G. We want to construct a 0-chain whose coboundary is c. Define the 0-chain a by the formula $a(y) = c(\tau_y)$. We claim that $(\delta a)(\gamma) = c(\gamma)$ for all paths γ. In fact,

$$(\delta a)(\gamma) \;=\; a(\gamma(1)) - a(\gamma(0)) \;=\; c(\tau_{\gamma(1)}) - c(\tau_{\gamma(0)}),$$

and $c(\tau_{\gamma(1)}) - c(\tau_{\gamma(0)}) = c(\gamma)$ since $\tau_{\gamma(1)} - \tau_{\gamma(0)} - \gamma$ is a 1-cycle, and c is assumed to vanish on 1-cycles.

To finish the proof, we must show that if $f\colon Z_1X \to G$ is a homomorphism that vanishes on B_1X, then f comes from some 1-cocycle c. Define the cochain c by the formula

$$c(\gamma) \;=\; f(\gamma + \tau_{\gamma(0)} - \tau_{\gamma(1)})$$

for any path γ. Equivalently, for any 1-chain γ, $c(\gamma) = f(\gamma - \Sigma m_i \tau_{y_i})$, where $\Sigma m_i y_i = \partial(\gamma)$. In particular, if γ is a 1-cycle, $c(\gamma) = f(\gamma)$, and if $\gamma = \partial\Gamma$, $c(\partial\Gamma) = f(\partial\Gamma) = 0$. So c is a 1-cocycle that maps to f. □

This proposition implies that $H^1(X;G)$ is isomorphic to the Čech group $H^1(\mathcal{U};G)$, provided $\mathcal{U}$ is a suitable cover of a nice space, so that Corollary 15.7 applies.

Exercise 16.11. When X is an open set in the plane, and $G = \mathbb{R}$, construct homomorphisms $H^0X \to H^0(X;\mathbb{R})$ and $H^1X \to H^1(X;\mathbb{R})$ from the De Rham groups to these cohomology groups, and show that they are isomorphisms.

There is also a Mayer–Vietoris theorem for these cohomology groups. If U' is an open subset of U there are natural restriction maps from $H^0(U;G)$ to $H^0(U';G)$ and from $H^1(U;G)$ to $H^1(U';G)$. If X is a union of two open sets U and V, there is a coboundary map

$$\delta\colon\; H^0(U \cap V;G) \;\to\; H^1(U \cup V;G).$$

To define this, given a 0-cocycle a on $U \cap V$, extend a to a 0-cochain $\tilde{a}$ on all of $U \cup V$ by defining $\tilde{a}$ to be zero on all points not in $U \cap V$. Then define the 1-cocycle $\delta(a)$ on $U \cup V$ whose value on a path γ is obtained by writing $\gamma = \gamma_1 + \gamma_2 + \tau$, where γ_1 is a 1-chain on U, γ_2 a 1-chain on V, and τ is a 1-boundary (see the proof of Lemma 10.2), and setting $\delta(a)(\gamma) = \tilde{a}(\partial\gamma_1)$.

Problem 16.12. Show that this definition is independent of choices, and verify that the resulting Mayer–Vietoris sequence

$$0 \longrightarrow H^0(U \cup V;G) \xrightarrow{+} H^0(U;G) \oplus H^0(V;G) \xrightarrow{-} H^0(U \cap V;G)$$
$$\xrightarrow{\delta} H^1(U \cup V;G) \xrightarrow{+} H^1(U;G) \oplus H^1(V;G) \xrightarrow{-} H^1(U \cap V;G)$$

is exact.

16d. G-Sets and Coverings

In this section and the next we describe two general constructions, which may help to put the constructions of §14a in context. The exercises verify the assertions made about these constructions.

Let $p\colon Y \to X = Y/G$ be a G-covering, without base point for the moment. Suppose T is any set, and we have a left action of G on T; we say that T is a *G-set*. Give T the discrete topology, so an action of G on T is a mapping from $G \times T$ to T satisfying properties (1) and (2) of §11c. The group G acts on the left on $Y \times T$ by the formula $g \cdot (y \times t) = g \cdot y \times g \cdot t$. Define Y_T to be the space of orbits:

$$Y_T = (Y \times T)/G.$$

Write $\langle y \times t \rangle$ in Y_T for the orbit containing $y \times t$. Let $p_T\colon Y_T \to X$ be the mapping that sends $\langle y \times t \rangle$ to $p(y)$.

Exercise 16.13. Show that the mapping $p_T\colon Y_T \to X$ is a covering map.

If T and T' are sets with G-actions, a *map* of G-sets is a function $\varphi\colon T \to T'$ such that $\varphi(g \cdot t) = g \cdot \varphi(t)$ for all t in T and g in G. A map of G-sets is an *isomorphism* if it is bijective, so there is an inverse mapping of G-sets from T' to T. A map $\varphi\colon T \to T'$ of G-sets determines a continuous mapping from $Y \times T$ to $Y \times T'$, taking $y \times t$ to $y \times \varphi(t)$. This is compatible with the actions of G, so it determines a continuous mapping from Y_T to $Y_{T'}$, taking $\langle y \times t \rangle$ to $\langle y \times \varphi(t) \rangle$, which commutes with the projections to X. If φ is an isomorphism, this is an isomorphism of coverings. Conversely, we have:

Exercise 16.14. Let $p\colon Y \to X = Y/G$ be a G-covering, with Y connected. Show that the two G-sets determine isomorphic coverings of X if and only if the G-sets are isomorphic. Show in fact that any continuous mapping f from Y_T to $Y_{T'}$ commuting with the projections to X comes from a map of G-sets from T to T'.

Exercise 16.15. If $T \to T'$ and $T' \to T''$ are maps of G-sets, show that the composite of $Y_T \to Y_{T'}$ and $Y_{T'} \to Y_{T''}$ is the mapping determined by the composite $T \to T''$.

Suppose the action of G on a set T is *transitive*: for any t_1 and t_2 in T, there is some g in G such that $g \cdot t_1 = t_2$. If a point t is chosen, let $H \subset G$ be the subgroup of elements of G that fix t, i.e., $H = \{g \in G: g \cdot t = t\}$. Then the G-set G/H of left cosets is isomorphic to the G-set T, by mapping the coset gH containing g to the point $g \cdot t$.

Exercise 16.16. Show that two transitive G-sets are isomorphic if and only if the corresponding subgroups of G are conjugate.

In Exercise 11.27 we saw how a subgroup H of G determines a covering $Y/H \to X$.

Exercise 16.17. If $T = G/H$ is a set of left cosets, show that the covering $Y_T \to X$ is isomorphic to the covering $Y/H \to X$.

Exercise 16.18. If T is a disjoint union of G-sets T_α, show that the covering $Y_T \to X$ is a disjoint union of the coverings $Y_{T_\alpha} \to X$.

An arbitrary G-set T is a disjoint union of its orbits T_α. By the preceding exercises the corresponding covering $Y_T \to X$ is a disjoint union of coverings of the form Y/H_α, for H_α subgroups of G.

Now suppose X is connected and locally path-connected, and has a universal covering space $\tilde{X} \to X$. Choose a point $\tilde{x}$ of $\tilde{X}$ lying over x in X. We have seen that, with these choices, the universal covering is a $\pi_1(X,x)$-covering. So every action of the fundamental group on a set T determines a covering $\tilde{X}_T \to X$. Combining what we have just proved with Proposition 13.23, we have:

Proposition 16.19. *Every covering of X is isomorphic to one obtained from the universal covering $\tilde{X} \to X$ by a left action of $\pi_1(X,x)$ on some set T. This covering is connected if and only if the action on T is transitive. Two such coverings are isomorphic if and only if the $\pi_1(X,x)$-sets are isomorphic.*

A left action of a group G on a set T is the same as a homomorphism of G to the group $\mathrm{Aut}(T)$ of permutations of T. Two such homomorphisms give isomorphic G-sets exactly when the homomorphisms are conjugate. In particular, taking $G=\pi_1(X,x)$ and $T=\{1, \ldots, n\}$, so $\mathrm{Aut}(T)$ is the symmetric group $\mathfrak{S}_n$ on n letters, we have a canonical bijection

$$\{n\text{-sheeted coverings of } X\}/\cong \;\leftrightarrow\; \mathrm{Hom}(\pi_1(X,x),\mathfrak{S}_n)/\text{conjugacy}.$$

16e. Coverings and Group Homomorphisms

Suppose $p\colon Y\to X$ is a G-covering, and $\psi\colon G\to G'$ is any homomorphism of groups. If we make G' into a left G-set, the construction of the preceding section can be used to construct a covering of X that is locally a product $X\times G'\to X$. We want to do this in such a way that this covering can be made into a G'-covering. The simplest way to make G' into a left G-set is by defining, for g in G and g' in G', $g\cdot g'$ to be $\psi(g)\cdot g'$, the latter using the group product in G'. However, we will use another, which will allow us to define a compatible left action of G' on the result. Define the left action of G on G' by

$$g\cdot g' \;=\; g'\cdot\psi(g^{-1}) \;=\; g'\cdot\psi(g)^{-1} \qquad \text{for} \quad g\in G \quad \text{and} \quad g'\in G'.$$

It is straightforward to check that this is a left action of G on G'. Now define $p(\psi)\colon Y(\psi)\to X$ to be the covering constructed from this left G-set. That is, $Y(\psi)$ is the quotient of $Y\times G'$ by the left action of G by $g\cdot(y\times g')=(g\cdot y\times g'\cdot\psi(g^{-1})$, with projection from $Y(\psi)$ to X determined by p on the first factor. Make G' act on $Y(\psi)$ by $g'\cdot\langle y\times h'\rangle=\langle y\times g'\cdot h'\rangle$.

Exercise 16.20. Show that $Y(\psi)\to X$ is a G'-covering of X. If Y has a base point y, $Y(\psi)$ gets the base point $\langle y\times e\rangle$. When p is the universal covering of X, this is the construction of Proposition 14.1.

Exercise 16.21. If $Y_\rho\to X$ is the G-covering determined by a homomorphism $\rho\colon \pi_1(X,x)\to G$, and $\psi\colon G\to G'$ is a homomorphism, show that the G'-covering $Y_\rho(\psi)\to X$ constructed from $Y_\rho\to X$ and ψ is isomorphic to the G'-covering $Y_{\psi\circ\rho}\to X$.

Exercise 16.22. Let $p\colon Y\to X$ be a G-covering, with Y connected. Let $\psi_1\colon G\to G'$ and $\psi_2\colon G\to G'$ be two homomorphisms. Show that, if base points are taken into account, the coverings $Y(\psi_1)$ and $Y(\psi_2)$ are isomorphic if and only if ψ_1 and ψ_2 are equal. Without base points, they are isomorphic if and only if ψ_1 and ψ_2 are conjugate, i.e., there is some g' in G' such that $\psi_2(g)=g'\cdot\psi_1(g)\cdot(g')^{-1}$ for all g in G.

Exercise 16.23. Show that if G is abelian, any G-covering of X is obtained from the universal abelian covering $\tilde{X}_{\text{abel}} \to X$ by a unique homomorphism ψ: $H_1X \to G$.

16f. G-Coverings and Cocycles

In §16d we constructed from a G-covering p: $Y \to X$ and a left action of G on a set T, a new covering p_T: $Y_T \to X$, where Y_T is the quotient of $(Y \times T)/G$, with G acting simultaneously on the left on Y and on T. This can also be described by a cocycle. Take a cocycle $\{g_{\alpha\beta}\}$ of transition functions from a trivialization of the covering $Y \to X$ over $\mathcal{U} = \{U_\alpha\}$, and use them to glue together the disjoint union of the $U_\alpha \times T$, identifying $x \times t$ in $U_\alpha \times T$ with $x \times g_{\beta\alpha}(x) \cdot t$ in $U_\beta \times T$ if x is in $U_\alpha \cap U_\beta$.

Exercise 16.24. Verify that this covering defined by a cocycle is isomorphic to that defined in §16d.

If ψ: $G \to G'$ is a homomorphism of groups, then ψ determines a map

$$\psi_*: H^1(\mathcal{U}; G) \to H^1(\mathcal{U}; G')$$

that takes a class represented by a cocycle $\{g_{\alpha\beta}\}$ to the class of $\{\psi(g_{\alpha\beta})\}$. (Exercise: Verify that this is well defined.) This means that any G-covering of X that is trivial over each U_α determines a G'-covering of X, also trivial over each U_α: if the G-covering is constructed by gluing the products $U_\alpha \times G$ with the transition functions $x \times g \mapsto x \times g \cdot g_{\alpha\beta}(x)$, then the corresponding G'-covering is constructed by gluing the products $U_\alpha \times G'$ with the transition functions $x \times g' \mapsto x \times g' \cdot \psi(g_{\alpha\beta}(x))$.

Problem 16.25. Show that this construction of a G'-covering from a G-covering agrees with the construction in §16d as the quotient of $Y \times G'$ by the action of G: $g \cdot (y \times g') = g \cdot y \times g' \cdot \psi(g^{-1})$: In particular, it is independent of choice of $\mathcal{U}$.

Exercise 16.26. A homomorphism ψ: $G \to G'$ determines a map from $\operatorname{Hom}(\pi_1(X,x), G)$ to $\operatorname{Hom}(\pi_1(X,x), G')$, taking ρ to $\psi \circ \rho$. Show that this is compatible with the bijections of Proposition 14.1.

In our definition of Čech cohomology $H^1(\mathcal{U}; G)$ there has been no topology on the group G, or, more accurately, we have equipped G with the discrete topology. Čech cohomology is also important when G is a group with a more interesting topology. One can define a set $H^1(\mathcal{U}; G)$ the same way, requiring the functions $g_{\alpha\beta}$ and h_α in the definition to be continuous maps from $U_{\alpha\beta}$ and U_α to G, respectively. From a cocycle $\{f_{\alpha\beta}\}$ one constructs in the same way a (principal) G-*bundle*, i.e., a space Y, on which G acts continuously, with a mapping $p\colon Y \to X$ so that $p(g \cdot y) = p(y)$, locally isomorphic to the projection $X \times G \to X$ (but in this product G is a topological space, not necessarily discrete).

If the topological group G acts continuously on the left on a space V, then a G-bundle determines a bundle with fiber V. The construction is as in the discrete case: take trivial coverings $U_\alpha \times V$, and glue them together using the identifications from $U_\alpha \times V$ to $U_\beta \times V$ over $U_\alpha \cap U_\beta$ by $x \times v \mapsto x \times g_{\beta\alpha}(x) \cdot v$; or one may construct the quotient $(Y \times V)/G$. The fundamental example is the case $G = \mathrm{GL}_n \mathbb{R}$ (with its usual topology as an open subset of $\mathbb{R}^{n^2}$), with $V = \mathbb{R}^n$, and the resulting bundle is called a *vector bundle* of rank n.

Problem 16.27. For $n = 1$, and $\mathcal{U}$ a covering of a plane open set U as above, show that the homomorphism from $H^1(\mathcal{U}; \{\pm 1\})$ to $H^1(\mathcal{U}; \mathrm{GL}_1 \mathbb{R})$ determined by the inclusion of $\{\pm 1\}$ in $\mathrm{GL}_1 \mathbb{R}$ is an isomorphism. Equivalently, giving a line bundle is equivalent to giving a two-sheeted covering. Prove this directly. Compute this group for U an annulus. Over a circle in the annulus you should find two line bundles, corresponding to a cylinder and a Moebius band.

Problem 16.28. Show how to make the set of all tangent vectors to the sphere $S^2 \subset \mathbb{R}^3$ into a vector bundle of rank 2. Generalize to construct the tangent bundle of an arbitrary manifold.

PART IX
TOPOLOGY OF SURFACES

The goal is to extend what we have done for open sets in the plane to open sets in surfaces, especially compact orientable surfaces. Some of this is straightforward generalization, valid for any surface, given basic notions about coordinate charts, but there are some new features that are special to the compact and orientable case.

We show that every compact orientable surface is homeomorphic to a sphere with g handles, at least under the assumption that the surface can be triangulated. Any such surface can be realized by taking a convex polygon with $4g$ sides, and making suitable identifications on the boundary. From this description it is routine to compute the fundamental group and first homology group.

In Chapter 18 we show that for any differentiable surface X, there is a canonical isomorphism between the De Rham group H^1X of closed mod exact 1-forms, and the dual group $\mathrm{Hom}(H_1X, \mathbb{R})$ to the homology group. This leads to a definition and calculation of an intersection pairing on the homology group H_1X.

Note: A surface is assumed to be connected unless otherwise stated. See Appendix D for foundational facts about surfaces.

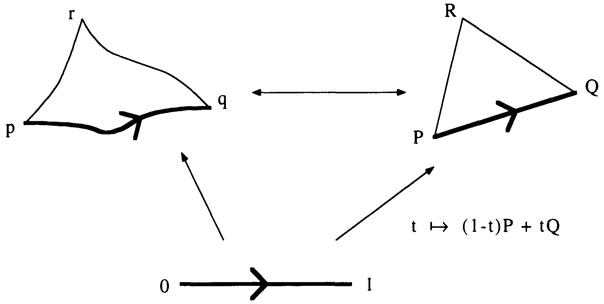
CHAPTER 17

The Topology of Surfaces

17a. Triangulation and Polygons with Sides Identified

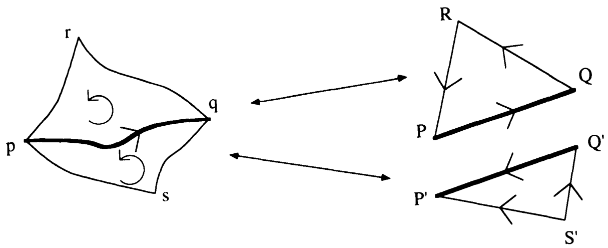
Our aim in this section is to show that a compact orientable surface is homeomorphic to a sphere with handles, under the assumption that the surface can be triangulated. The idea is to "flatten" it out, to realize the surface as a polygon in the plane, with certain identifications on the boundary edges, and then, by cutting and pasting, to simplify these realizations until we can recognize the result.

A *triangulation* of a compact surface X has a finite number f of *faces*, each of which is a subset of X homeomorphic to an ordinary (closed) triangle in the plane, with three *edges* homeomorphic to closed intervals, and three *vertices* that are points. Two edges can meet only at a vertex, and two faces meet only at one vertex or exactly along a common edge; in the latter case the faces are called *adjacent*. There are many choices for such homeomorphisms between plane triangles and faces of the triangulation. In order to be able to compare the homeomorphisms on adjacent faces in X, we first choose a homeomorphism of each edge on X with the interval $[0, 1]$. We can then find homeomorphisms of each face on X with a plane triangle so that, on the edges, the maps are those determined by these homeomorphisms:

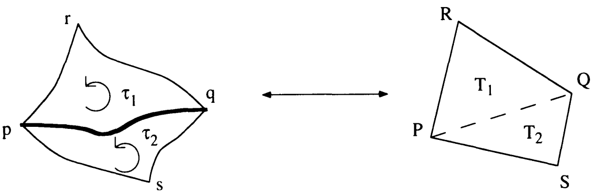


If two triangles correspond to adjacent faces, this means that the resulting homeomorphism between corresponding sides $\overline{PQ}$ and $\overline{P'Q'}$ is given by the "affine" map $(1-t)P + tQ \mapsto (1-t)P' + tQ'$. When we identify sides of polygons, it will always be by such affine homeomorphisms, so there will be no ambiguity in the results.

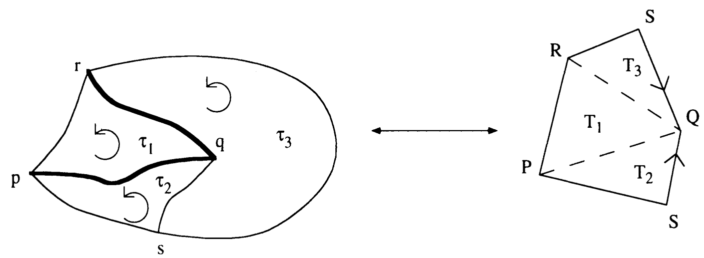
An orientation of X determines a counterclockwise direction around the boundary of each triangle, with adjacent triangles determining opposite directions along their common edge:



Choose one face τ_1 in the triangulation, and choose a homeomorphism of τ_1 with a plane triangle $\Pi_1 = T_1$, with counterclockwise orientation. Choose a face τ_2 adjacent to τ_1, and extend the homeomorphism along their common edge to a homeomorphism of τ_2 with a triangle T_2 adjacent to T_1:

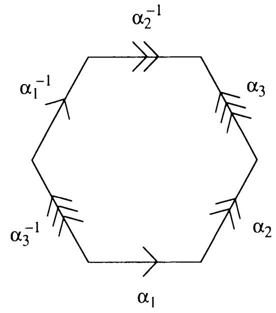


Choose T_2 so the union of T_1 and T_2 is a convex quadrilateral Π_2. This gives a homeomorphism from $\tau_1 \cup \tau_2$ to Π_2. Now choose τ_3 adjacent to τ_1 or τ_2 (if τ_3 is adjacent to both, choose one of them arbitrarily), and extend the map from the common edge of τ_3 and the (chosen) τ_1 or τ_2 to a homeomorphism from τ_3 to a triangle T_3 adjacent to the corresponding side of Π_2. Let Π_3 be the resulting five-sided polygon, which we can take to be convex:



The arrows indicate the identification made in the map from the figure on the right onto that on the left. Continue in this way, until all f of the faces have been used. We then have a convex polygon $\Pi_f = \Pi$ with $f + 2$ sides. Each side of Π will correspond to a common edge of two faces on X, and each edge occurring this way will correspond to two such sides of Π. Thus the sides of Π will be paired off, and we have a continuous map from Π to X that realizes X as a quotient space of Π, obtained by identifying corresponding points on corresponding sides. Note that, when traveling around Π in a counterclockwise direction, one travels along two corresponding sides in opposite directions.

This realization of X as an identification space of a polygon is half way to our goal. The data used in constructing X from the polygon can be encoded by taking an alphabet with m letters, with $m = {}^1\!/_2(f + 2)$, and writing down a sequence of $2m$ symbols, each of which is a letter α in the alphabet, or its "inverse" α^{-1}, with each of these symbols occurring just once. For example, the space constructed by making the identifications indicated in the following diagram



can be described by the code $\alpha_1\alpha_2\alpha_3\alpha_2{}^{-1}\alpha_1{}^{-1}\alpha_3{}^{-1}$.

Exercise 17.1. Show that this code represents a torus.

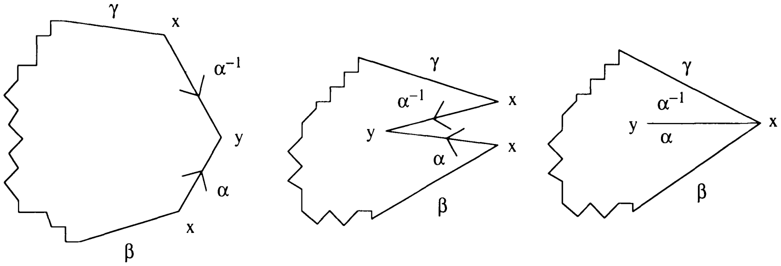
Exercise 17.2. If the number of sides is four, show that X is homeomorphic to a sphere or a torus.

There is a good deal of arbitrariness in the choice of polygon and its code, even for a given triangulation and ordered choice of faces of X. The polygon can certainly be replaced by any other convex polygon with the same number of sides. (For simplicity, one may use a regular polygon.) Of course, even when the polygon and the identifications are fixed, the choice of alphabet is arbitrary, as is the choice of which of a pair of corresponding sides is written as a letter α of the alphabet and which as α^{-1}. In addition, the place where one starts listing the sides is arbitrary, and the code can therefore by cyclically permuted. Note that the identifications are made by identifying corresponding points on corresponding sides; some vertices of Π may become identified by this, and others may not.

17b. Classification of Compact Oriented Surfaces

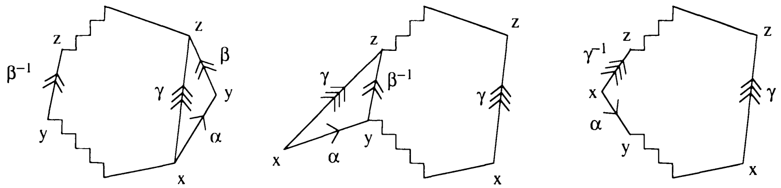
For the rest of the argument we do not need the triangulation, but only this description as a polygon with sides identified according to a code. We want to prescribe some rules for simplifying a polygonal presentation of a surface. We proceed in several steps. By Exercise 17.2 we can assume the polygon has $2m$ sides, with $m \geq 3$.

Step 1. If a letter α and its inverse α^{-1} occur successively in the code, they can both be omitted. That is, X is homeomorphic to the surface obtained from a convex polygon with $2m-2$ sides, using the same code in the same order but with α and α^{-1} omitted. This can be seen from the pictures



The point is that if one first carries out the identification of α with α^{-1}, one gets a space homeomorphic to a convex polygon, so that the remaining identifications are prescribed by the code with α and α^{-1} omitted.

Step 2. It can be assumed that all the vertices of the polygon map to the same point of X. To show this, it suffices, if not all the vertices have the same image, to show how to increase by one the number of vertices mapping to a given point x, without changing the total number of vertices. There is a side, say labeled α, that joins two vertices, one of which is mapped to x and the other to a point y not equal to x. Let β be the side of Π adjacent to α at the point mapping to y. Draw the diagonal γ as shown on the polygon, and cut off the triangle formed by α, β, and γ, and attach it to the polygon by identifying β with β^{-1}:



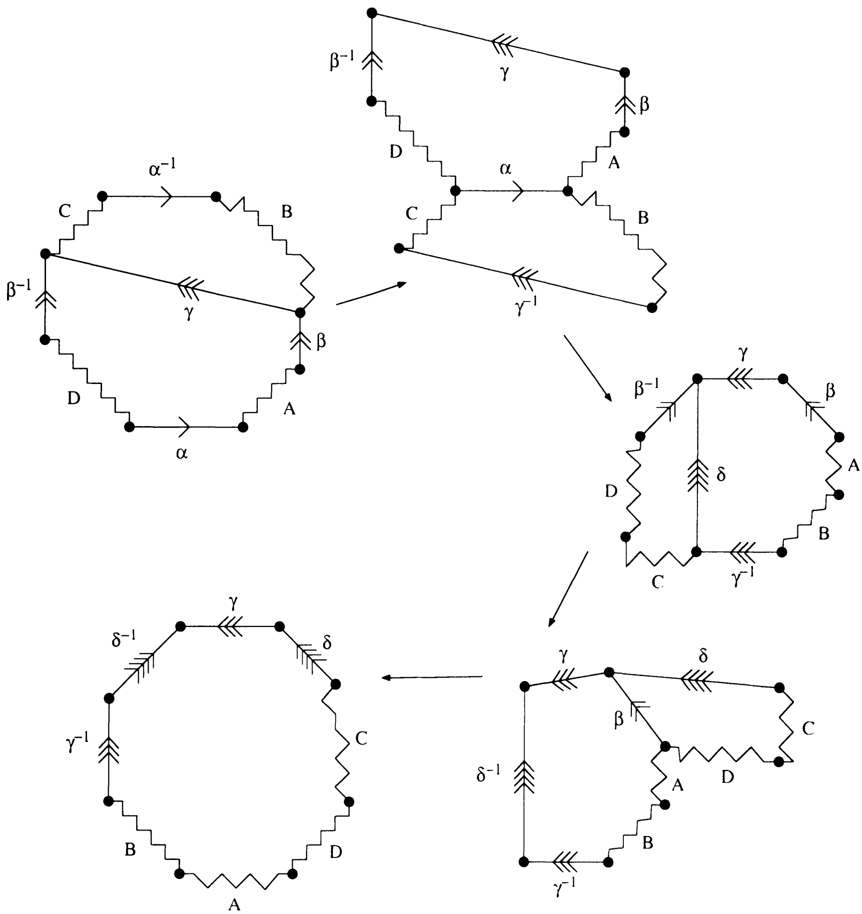
The surface X is the identification space of this new polygon, and it has the same number of sides, but one more vertex mapping to x.

Step 3. By steps 1 and 2 we may assume that all vertices map to the same point x in X, and that no edge is adjacent to its inverse. We claim next that for any edge α there must be an edge β lying between α and α^{-1} so that β^{-1} lies between α^{-1} and α; that is, in a code, after cyclically permuting if necessary, they occur in the order

$$\ldots \alpha \ldots \beta \ldots \alpha^{-1} \ldots \beta^{-1} \ldots \quad .$$

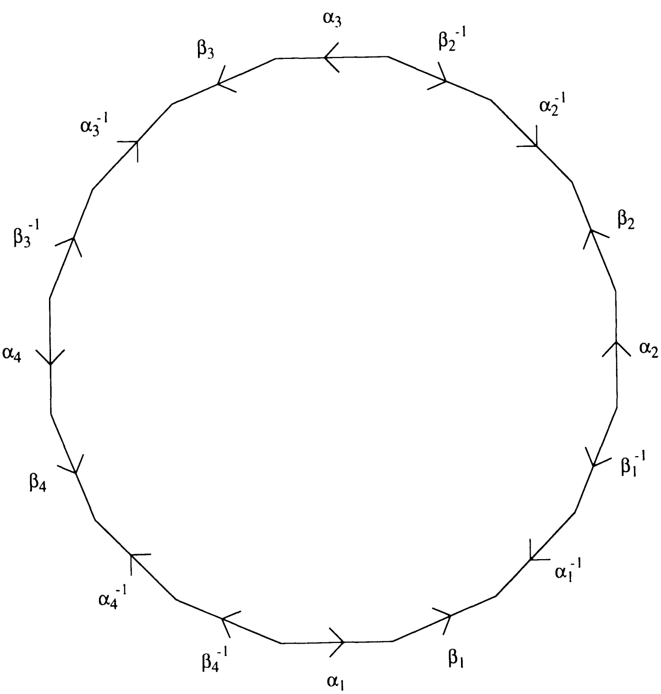
If not, we could construct the identification space X by first identifying α with α^{-1}, and then separately doing all the identification prescribed by edges lying (counterclockwise) between α and α^{-1}, and those prescribed by edges lying (counterclockwise) between α^{-1} and α. The two endpoints of α never get identified in this process, contradicting the assumption that all the vertices map to the same point of X.

Now we simplify the code as follows: Choose an edge α, and take an edge β so that the edges α, β, α^{-1}, and β^{-1} occur in this order, but perhaps with other edges in between some of these (denoted A, B, C, and D in the following diagram). Perform the following sequence of cutting and pasting moves:

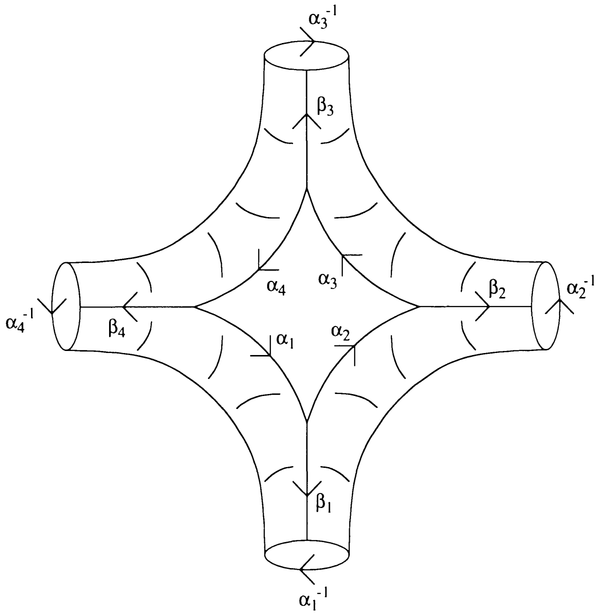


This gives a new polygon with the same properties and the same number of vertices, which also represents X, but now it has a successive sequence $\delta \cdot \gamma \cdot \delta^{-1} \cdot \gamma^{-1}$. In addition, if other such sequences occurred elsewhere in the polygon (i.e., in one of the parts labeled A, B, C, or D), such sequences are not disturbed by this procedure. So we may continue this process, until we have represented X as a polygon with sides identified according to a code

(17.3)
$$\alpha_1 \cdot \beta_1 \cdot \alpha_1^{-1} \cdot \beta_1^{-1} \cdot \alpha_2 \cdot \beta_2 \cdot \alpha_2^{-1} \cdot \beta_2^{-1} \cdot \ldots \cdot \alpha_g \cdot \beta_g \cdot \alpha_g^{-1} \cdot \beta_g^{-1}.$$

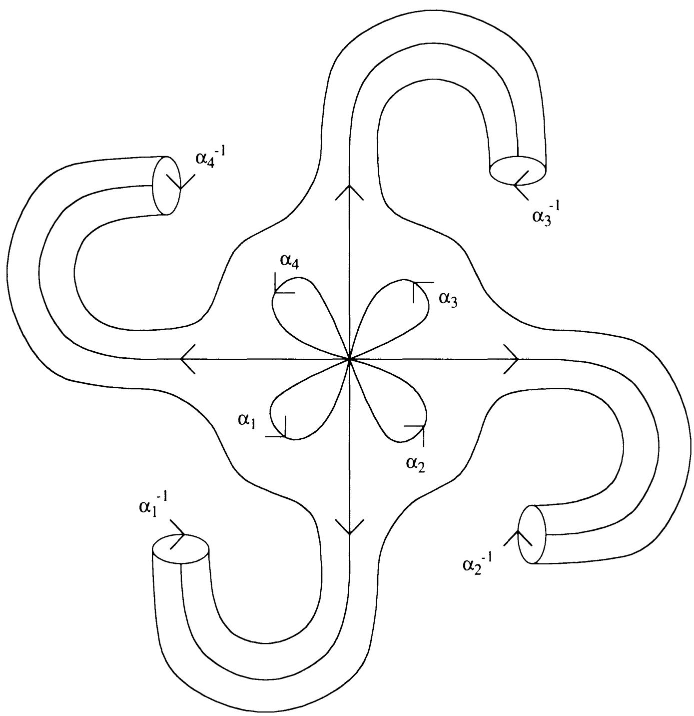


This is called a *normal form*. From this one can see directly that the identification space is a sphere with g handles. First make the identification of each β_i with β_i^{-1}:



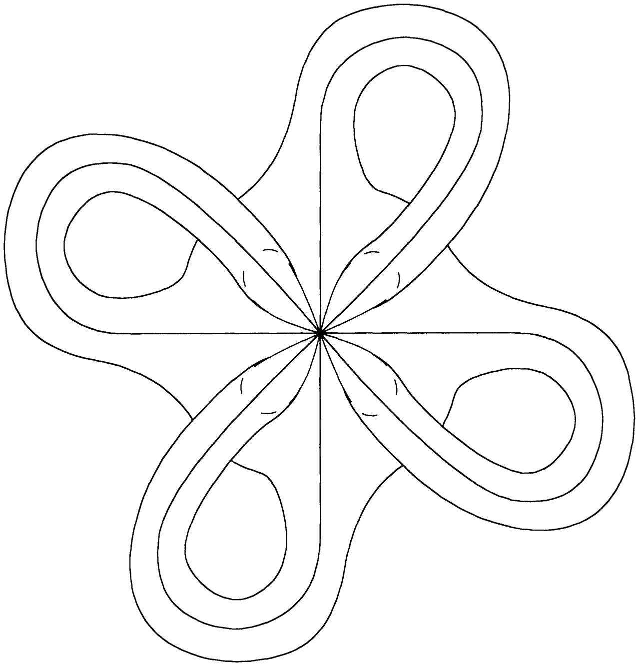
α_3^{-1}
β_3
α_4
α_3
β_2
α_2^{-1}
α_4^{-1}
β_4
α_1
α_2
β_1
α_1^{-1}

or



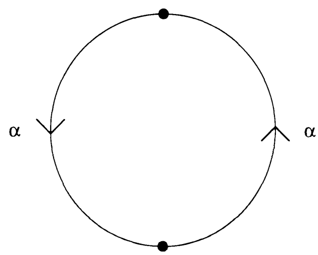
α_4^{-1}
α_3^{-1}
α_4
α_3
α_1
α_2
α_1^{-1}
α_2^{-1}

Then identify each α_i with α_i^{-1}:



Theorem 17.4. *If X is a (triangulable) compact orientable surface, then X is homeomorphic to a sphere with g handles for some nonnegative integer g.*

A similar procedure leads to a normal form for nonorientable compact surfaces. The projective plane can be realized by the code $\alpha \cdot \alpha$, identifying the opposite sides of a "two-sided" polygon



Problem 17.5. Show that a (triangulable) compact nonorientable sur-

face has a normal form given by a code $\alpha_1 \cdot \alpha_1 \cdot \alpha_2 \cdot \alpha_2 \cdot \ldots \cdot \alpha_h \cdot \alpha_h$, for some positive integer h.

For a discussion of triangulation of arbitrary compact surfaces, together with a more modern proof of Theorem 17.4, see Armstrong (1983).

17c. The Fundamental Group of a Surface

The normal form representation of a compact orientable surface X as a $4g$-sided polygon Π with sides identified according to the code (17.3) can be used to compute the fundamental group of X. Let x be the point in X that is the image of the vertices in Π, and let α_i and β_i be the loops in X that are the images of the corresponding sides of Π.

Let F_{2g} be the free group on $2g$ generators $a_1, b_1, \ldots, a_g, b_g$. There is a homomorphism of groups from F_{2g} to the fundamental group $\pi_1(X,x)$ that takes a_i to α_i and b_i to β_i for $1 \le i \le g$. We claim first that the element

$$c_g = a_1 \cdot b_1 \cdot a_1^{-1} \cdot b_1^{-1} \cdot a_2 \cdot b_2 \cdot a_2^{-1} \cdot b_2^{-1} \cdot \ldots \cdot a_g \cdot b_g \cdot a_g^{-1} \cdot b_g^{-1}$$

in F_{2g} maps to the identity element of $\pi_1(X,x)$ by this homomorphism. In fact, the product path

$$\alpha_1 \cdot \beta_1 \cdot \alpha_1^{-1} \cdot \beta_1^{-1} \cdot \alpha_2 \cdot \beta_2 \cdot \alpha_2^{-1} \cdot \beta_2^{-1} \cdot \ldots \cdot \alpha_g \cdot \beta_g \cdot \alpha_g^{-1} \cdot \beta_g^{-1}$$

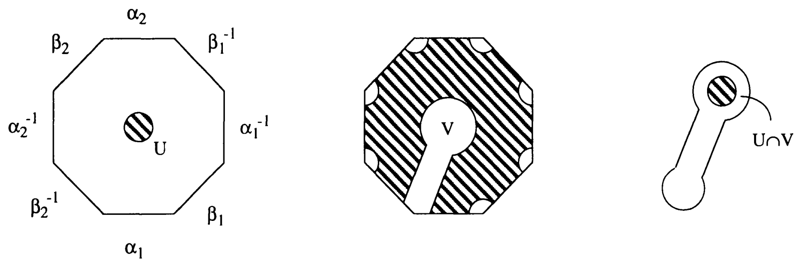
is homotopic to the constant path at x, since the corresponding path around the sides of Π is homotopic to a constant path in Π (since Π is convex), and composition with the continuous map from Π to X gives a homotopy in X.

Let N_g be the least normal subgroup of F_{2g} containing c_g, i.e., N_g is the subgroup generated by all elements of the form $u \cdot c_g \cdot u^{-1}$ for all u in F_{2g}. From what we have just seen, N_g is in the kernel of the homomorphism from F_{2g} to $\pi_1(X,x)$, so we have a homomorphism

$$F_{2g}/N_g \to \pi_1(X,x).$$

Proposition 17.6. *This homomorphism $F_{2g}/N_g \to \pi_1(X,x)$ is an isomorphism.*

Proof. We will apply the Van Kampen theorem. Let U be the image in X of the complement of a small disk in the middle of Π, and let V be the image in X of an open set that contains this disk as shown:



Let K be the image of the boundary of Π in X. This K is a graph that consists of $2g$ loops at x, so $\pi_1(K,x)$ is the free group on generators $a_1, b_1, \ldots, a_g, b_g$, i.e., $\pi_1(K,x) = F_{2g}$. By radial projection from the center of the disk, one sees that K is a deformation retract of U. We therefore have an isomorphism

$$\pi_1(U,x) \cong \pi_1(K,x),$$

so $\pi_1(U,x)$ is also the free group F_{2g} on the same generators. Now V is homeomorphic to a disk, so $\pi_1(V,x) = \{e_x\}$ is trivial, and $U \cap V$ has a circle for a deformation retract, so $\pi_1(U \cap V, x) = \mathbb{Z}$. The inclusion of $U \cap V$ in U takes a generator of $\pi_1(U \cap V, x)$ to the element $c_g = a_1 \cdot b_1 \cdot a_1^{-1} \cdot \ldots \cdot b_g^{-1}$. By Van Kampen's theorem, for any group G, to give a homomorphism from $\pi_1(X,x)$ to G is the same as giving a homomorphism from $\pi_1(U,x) = F_{2g}$ to G (and a homomorphism from $\pi_1(V,x) = \{e\}$ to G) in such a way that the composite

$$\mathbb{Z} = \pi_1(U \cap V, x) \to \pi_1(U,x) \to G$$

takes a generator of $\mathbb{Z}$ to the identity element of G. This means precisely that one has a homomorphism from F_{2g} to G such that c_g maps to the identity, or equivalently, that N_g maps to the identity. That is,

$$\mathrm{Hom}(\pi_1(X,x), G) = \mathrm{Hom}(F_{2g}/N_g, G).$$

This implies that the map from F_{2g}/N_g to $\pi_1(X,x)$ is an isomorphism, cf. Exercise 14.5. (See also Exercise 14.8, which applies to this situation directly.) □

Corollary 17.7. *The first homology group H_1X is a free abelian group of rank $2g$, with basis the image of the loops $\alpha_1, \beta_1, \ldots, \alpha_g, \beta_g$.*

Proof. Let A_{2g} be the free abelian group on $2g$ generators. The canonical map from F_{2g} to A_{2g} maps N_g to 0, since c_g is in the commutator subgroup. This determines a homomorphism from

$\pi_1(X,x) = F_{2g}/N_g$ to A_{2g}, and therefore a homomorphism

$$H_1(X) = \pi_1(X,x)/[\pi_1(X,x), \pi_1(X,x)] \to A_{2g}.$$

Since the images of $a_1, b_1, \ldots, a_g, b_g$ generate $H_1(X)$, and their images in A_{2g} are linearly independent, it follows that they are linearly independent in $H_1(X)$ and that this map is an isomorphism. □

In particular, since the rank of a free abelian group is an invariant of the group, this shows that the number g of Theorem 17.4 is independent of choices, and depends only on the topology of X. It is called the *genus* of the surface. Together with what we saw in Chapter 8, this shows that for any triangulation of X, the number v of vertices, e of edges, and f of faces satisfies the Euler equation

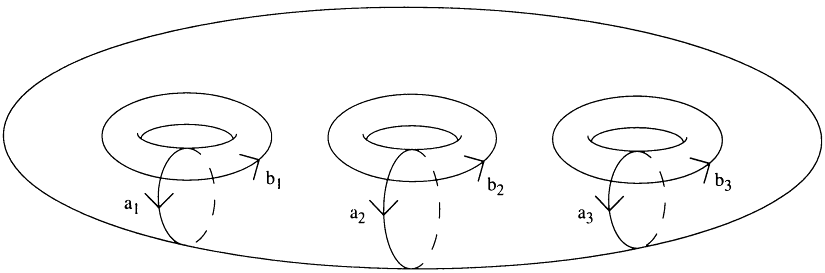
$$v - e + f = 2 - 2g.$$

Problem 17.8. For an nonorientable compact surface with normal form $\alpha_1 \cdot \alpha_1 \cdot \alpha_2 \cdot \alpha_2 \cdot \ldots \cdot \alpha_h \cdot \alpha_h$ as in Problem 17.5, show that the fundamental group is the quotient of the free group F_h on h generators $a_1, \ldots, a_h$ by the least normal subgroup that contains the element $a_1^2 \cdot a_2^2 \cdot \ldots \cdot a_h^2$. Show that the first homology group is isomorphic to $\mathbb{Z}^{\oplus(h-1)} \oplus (\mathbb{Z}/2\mathbb{Z})$. In particular, the number h is independent of all choices.

Exercise 17.9. Prove that for any triangulation of a compact non-orientable surface, with h as above, the number v of vertices, e of edges, and f of faces satisfies the Euler equation

$$v - e + f = 2 - h.$$

For homology there is no need to use paths with the same base points. It will be convenient to change the paths α_i and β_i to paths that we denote by a_i and b_i as shown:



Exercise 17.10. Show that the loops α_i and a_i define the same classes

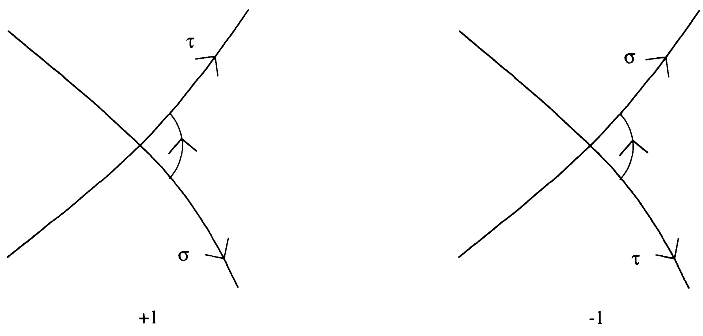
in $H_1(X)$, and similarly for β_i and b_i. In particular, the $2g$ classes determined by $a_1, b_1, \ldots, a_g, b_g$ form a free basis for $H_1(X)$.

The polygonal normal form of a surface can also be used to describe its universal covering space. When $g = 0$, of course, X is a sphere, so simply connected. When $g = 1$, X is a torus, which we have seen has the plane $\mathbb{R}^2$ as its universal covering. One can construct this universal covering from its representation as a rectangle with sides identified, by pasting the sides together but without identifying opposite sides. For $g \geq 2$, the universal covering can be realized in a similar way by gluing together copies of the polygon Π, but one cannot do this in a metrical way in the ordinary plane. However, in a hyperbolic plane this can be done, and one can see that the universal covering can be realized as a hyperbolic plane. The hyperbolic plane is homeomorphic to an open disk (or to $\mathbb{R}^2$), so from this one sees that for all $g \geq 2$ the universal covering is an open disk. For more on the hyperbolic plane, see Hilbert and Cohn-Vossen (1952).

Exercise 17.11. Give another proof of the results of this section by computing the fundamental group of the complement of $2g$ disjoint disks in a two-sphere, and then showing what happens when one sews g handles (each homeomorphic to $S^1 \times [0, 1]$) onto pairs of the boundaries of these disks.

Problem 17.12. Compute the fundamental group and first homology group of the space obtained from a surface X of genus g by removing n disjoint disks or points.

There is another important operation on the first homology group H_1X of an oriented compact surface, an *intersection pairing*, that assigns a number $\langle \sigma, \tau \rangle$ to two classes σ and τ in H_1X. If the classes are represented by loops that meet each other transversally in a finite number of points, this number is the sum of the numbers ± 1 assigned to each point of intersection, with the number ± 1 assigned according to the following picture:



The number is $+1$ if the direction from σ to τ is counterclockwise, and -1 if it is clockwise. (Note that this notion depends on having an orientation for X.) What takes some proof is showing that this gives a well-defined number: that any two classes in H_1X have representatives that meet transversally (which is not difficult), and then that the number one gets is independent of choices (which is more difficult). Rather than doing this directly, we will construct this intersection pairing by another procedure in the next chapter, by relating it to the wedge product of 1-forms.

CHAPTER 18

Cohomology on Surfaces

18a. 1-Forms and Homology

On any $\mathscr{C}^\infty$ surface X, just as on an open set in the plane, we have a notion of $\mathscr{C}^\infty$ functions, 1-forms, and 2-forms ω, and we have linear maps d, that take a function f to a 1-form df, and a 1-form ω to a 2-form $d\omega$. (See Appendix D3 for definitions and basic properties.) Therefore we have a notion of a 1-form ω being *closed* ($d\omega = 0$) or *exact* ($\omega = df$ for some f). All exact forms are closed, so we can define the first De Rham cohomology group as before:

$$H^1X = \{\text{closed 1-forms on } X\}/\{\text{exact 1-forms on } X\}.$$

Just as in the case of open sets in the plane, if $\gamma\colon [a,b] \to X$ is a continuous path, and ω is a closed 1-form on X, we can define an integral $\int_\gamma \omega$. As before, if γ is differentiable, this can be done by calculus: $\int_\gamma \omega = \int_a^b \gamma^*\omega$. In general, as in the plane, we can subdivide the interval by $a = t_0 \leq t_1 \leq \ldots \leq t_n = b$, so that $\gamma([t_{i-1}, t_i])$ is contained in an open set U_i on which $\omega = df_i$, and then

$$\int_\gamma \omega = \sum_{i=1}^{n} f_i(\gamma(t_i)) - f_i(\gamma(t_{i-1})).$$

The same argument as in §9b shows that this is independent of choices. As before, this extends from integrals over paths to integrals over 1-chains, and the same argument shows that the integral over a boundary is zero, and that the integral of df over a cycle is zero. It follows that we again have a canonical map

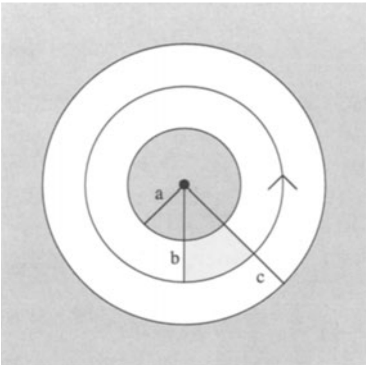
$$H^1X \to \mathrm{Hom}(H_1X, \mathbb{R})$$

that takes the class of ω to the homomorphism $[\gamma] \mapsto \int_\gamma \omega$.

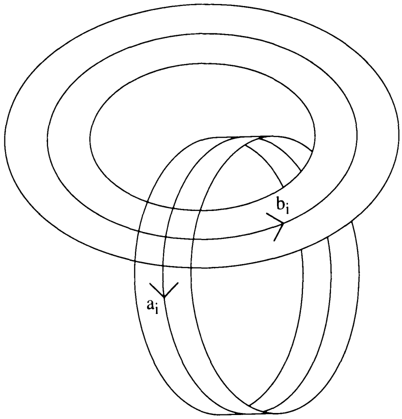
We want to calculate H^1X when X is a compact oriented surface. In particular, we want to show that this canonical map is an isomorphism. This could be done by the Mayer–Vietoris argument (which generalizes from open sets in the plane), but we will do it by explicitly constructing a basis, which will be useful later.

We take a model of X as in §17c, with a basis for H_1X given by the loops $a_1, b_1, \ldots, a_g, b_g$ as indicated in the picture there. For each loop a_i we will construct a closed 1-form α_i, and for each b_i a closed 1-form β_i, and we will show that the classes of these $2g$ 1-forms form a basis of H^1X. The forms α_i and β_i will depend on several choices, but we will see that their classes in H^1X depend only on the homology classes of the $2g$ loops $a_1, b_1, \ldots, a_g, b_g$.

For each of these loops we can find an open set V containing it that is diffeomorphic to an annulus $U = \{(x, y): a^2 < x^2 + y^2 < c^2\}$, with the loop corresponding to a counterclockwise circle around the middle of the annulus, i.e., to the path $t \mapsto (b\cos(2\pi t), b\sin(2\pi t))$, $0 \le t \le 1$, for some $a < b < c$:



We take these neighborhoods V to be disjoint, except for those containing the same a_i and b_i, which can be taken to intersect in a set diffeomorphic to an open rectangle.



Consider one such loop, either a_i or b_i, and fix such a diffeomorphism φ from a neighborhood V of the loop to such an annulus U. From Exercise B.14 we can find a $\mathscr{C}^\infty$ function ψ on the plane that is identically 1 outside the big circle and identically 0 inside the small circle. In fact, for any ε with $0 < \varepsilon < {}^1\!/_2(c-a)$, we can find ψ so that

$$\psi(x, y) = \begin{cases} 1 & \text{if } x^2+y^2 \geq (c-\varepsilon)^2, \\ 0 & \text{if } x^2+y^2 \leq (a+\varepsilon)^2. \end{cases}$$

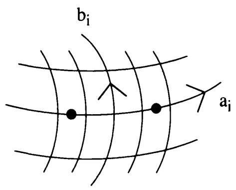
The function $\psi \circ \varphi$ is a $\mathscr{C}^\infty$ function on V that is constant on a neighborhood of the boundary of V. Its differential $d(\psi \circ \varphi)$ is therefore a 1-form that is identically zero on a neighborhood of the boundary of V. We can therefore define a $\mathscr{C}^\infty$ 1-form on all of X by defining it to be $d(\psi \circ \varphi)$ on V, and identically 0 outside V. This is certainly a closed 1-form. In fact, its restriction to V is exact, and its restriction to an open set V' with $V \cup V' = X$ is zero. (Note that the form is *not* exact on X, however, as follows for example from the following lemma.) This is the 1-form we wanted to construct. We denote it by α_i if the given loop is a_i, and β_i if the given loop is b_i.

Lemma 18.1. *These 1-forms α_i and β_i have integrals:*

(i) $\displaystyle\int_{b_j} \beta_i = 0$ *for all i and j, and* $\displaystyle\int_{a_j} \beta_i = \begin{cases} 1 & \textit{if } i=j, \\ 0 & \textit{if } i \neq j; \end{cases}$

(ii) $\displaystyle\int_{a_j} \alpha_i = 0$ *for all i and j, and* $\displaystyle\int_{b_j} \alpha_i = \begin{cases} -1 & \textit{if } i=j, \\ 0 & \textit{if } i \neq j. \end{cases}$

Proof. This is a direct calculation. All of the assertions that integrals vanish follow from the fact that they are integrals of 1-forms over a loop, and these forms are exact on an open set containing the loop. For the integral of β_i along a_i, cut a_i into two pieces, one inside the annulus as shown, the other outside.



For the piece inside the annulus, β_i is the differential of a function whose value at the initial point is 0 and at the end point is 1, so the integral of β_i along this piece is 1. For the second, β_i is identically zero. So the integral is $1 + 0 = 1$, as asserted. The argument is similar for the integral of α_i along b_i, but this time the function is 1 at the initial point and 0 at the final point of the part of the path in the annulus, so the integral is -1. □

Proposition 18.2. (1) *The canonical map $H^1X \to Hom(H_1X, \mathbb{R})$ is an isomorphism.*

(2) *The classes of the 1-forms $\alpha_1, \beta_1, \ldots, \alpha_g, \beta_g$ form a basis for H^1X.*

Proof. The same argument as in the plane shows that this canonical map is one-to-one. For if ω is a closed 1-form and $\int_\gamma \omega = 0$ for all closed paths γ, then one can define a function f on X by fixing a point P_0 and defining $f(P)$ to be the integral of ω along any path from P_0 to P. The assumptions make this a well-defined function, and the proof that $df = \omega$ is the same as for open sets in the plane (see Proposition 1.8); indeed, the assertion $df = \omega$ is a local assertion, so it can be verified on coordinate neighborhoods.

To see that the canonical map is subjective, a homomorphism h: $H_1X \to \mathbb{R}$ is determined by its values on a basis, so suppose $h([a_j]) = r_j$ and $h([b_j]) = s_j$ for some numbers r_j and s_j, $1 \leq j \leq g$. By the lemma, the closed 1-form $\omega = \Sigma_{i=1}^{g}(r_i \beta_i - s_i \alpha_i)$ has $\int_{a_j}\omega = r_j$ and $\int_{b_j}\omega = s_j$, which means that the canonical map takes ω to h. □

Exercise 18.3. Show that for closed 1-forms ω on the surface X one has a "module of periods" story as in Chapter 9: if the integrals of

ω along the basic loops are known, all integrals are determined. Show in fact that if $\int_{a_j}\omega = r_j$ and $\int_{b_j}\omega = s_j$, then for any closed 1-chain γ,

$$\int_\gamma \omega = \sum_{i=1}^{g} (m_i r_i + n_i s_i),$$

for some integers m_i and n_i.

Note in particular that the classes $[\alpha_i]$ and $[\beta_i]$ of α_i and β_i in H^1X are determined by their integrals, so are independent of the choices we made in defining them. They depend in fact only on the choice of the basis for H_1X determined by the a_i and b_i.

18b. Integrals of 2-Forms

Once an orientation of our surface is chosen, one can integrate $\mathscr{C}^\infty$ 2-forms. The integral of the 2-form ν on X will be a real number, denoted $\iint_X \nu$. More generally, if X is any oriented surface, and ν is a 2-form with compact support (i.e., which is identically zero outside a compact subset of X), one can define $\iint_X \nu$ by the following procedure. Choose an atlas of coordinate charts $\varphi_\alpha\colon U_\alpha \to X$, say with each U_α an open rectangle in the plane, with all charts compatible with the given orientation, and satisfying the condition that any point of X has a neighborhood meeting only finitely many $\varphi_\alpha(U_\alpha)$ (see Appendix A4 and Lemma 24.10).

To give a 2-form ν on X is the same as giving a 2-form $\nu_\alpha\, dx\, dy$ on U_α (which the same as prescribing a function ν_α on U_α), such that these 2-forms are compatible under changes of coordinates. First suppose ν is a 2-form that is zero except on a compact subset of one of the open sets $\varphi_\alpha(U_\alpha)$. Then one can define the integral of ν by

$$\iint_X \nu = \iint_{U_\alpha} \nu_\alpha\, dx\, dy,$$

where the integral on the right is the ordinary Riemann integral of a $\mathscr{C}^\infty$ (or continuous) function ν_α on a rectangle. In general, choose a partition of unity $\{\psi_\alpha\}$ so that the closure of the support of each ψ_α is contained in $\varphi_\alpha(U_\alpha)$. Define the integral of ν to be the sum of the integrals of the $\psi_\alpha \nu$, i.e.,

$$\iint_X \nu = \sum_\alpha \left(\iint_X \psi_\alpha \nu \right),$$

where the integral of each $\psi_\alpha \nu$ is defined by the preceding case, since it is zero except on $\varphi_\alpha(U_\alpha)$. Note by the compactness of the support of ν that only finitely many $\psi_\alpha \nu$ can be nonzero, so this sum is finite.

Exercise 18.4. (a) Verify that this definition is independent of the choice of partition of unity and the choice of coordinate charts. (b) Verify that the integral is linear: $\iint_X (r_1\nu_1 + r_2\nu_2) = \iint_X r_1\nu_1 + \iint_X r_2\nu_2$ for 2-forms ν_1 and ν_2 and real numbers r_1 and r_2. (c) Show that, if the other orientation is chosen, then the integral changes sign.

We need the following version of the Green–Stokes theorem for a surface:

Proposition 18.5. *If $\nu = d\omega$, where ω is a 1-form with compact support on an oriented surface X, then $\iint_X \nu = 0$.*

Proof. Choosing an atlas and partition of unity as in the definition, since $\omega = \Sigma_\alpha \psi_\alpha \omega$,

$$\iint_X \nu = \iint_X d\omega = \iint_X d\Big(\sum_\alpha \psi_\alpha\omega\Big) = \sum_\alpha \iint_X d(\psi_\alpha\omega).$$

Now $\psi_\alpha\omega$ is a 1-form with compact support on $\varphi_\alpha(U_\alpha)$, which corresponds to a 1-form μ_α with compact support on the open rectangle U_α. By definition, $\iint_X d(\psi_\alpha\omega) = \iint_{U_\alpha} d(\mu_\alpha)$, so it suffices to prove that this last integral is zero. But μ_α has compact support, so it vanishes on the boundary of the rectangle. By Green's theorem for a rectangle (Lemma 1.11),

$$\iint_{U_\alpha} d(\mu_\alpha) = \int_{\partial U_\alpha} \mu_\alpha = 0. \qquad \square$$

18c. Wedges and the Intersection Pairing

If X is a compact oriented surface, and ω and μ are 1-forms on X, we can define a real number (ω, μ) by the formula

$$(\omega, \mu) = \iint_X \omega \wedge \mu.$$

where $\omega \wedge \mu$ is the wedge product (Appendix D3). From basic properties of the wedge product (see Exercise D.7) we have:

(i) $(r_1\omega_1 + r_2\omega_2, \mu) = r_1(\omega_1, \mu) + r_2(\omega_2, \mu)$ for 1-forms ω_1, ω_2, and μ, and real numbers r_1 and r_2;
(ii) $(\omega, \mu) = -(\mu, \omega)$ for any 1-forms ω and μ; and
(iii) $(df, \mu) = 0$ for a function f and a closed 1-form μ.

Condition (i) says that (ω, μ) is linear in the first factor. Similarly, or using (ii), it is linear in the second, i.e., it is a *bilinear pairing*. Condition (ii) says that this pairing is *skew-symmetric*. Condition (iii), together with (ii), says that $(\omega, \mu) = 0$ if either ω or μ is exact. It follows that we can define a mapping

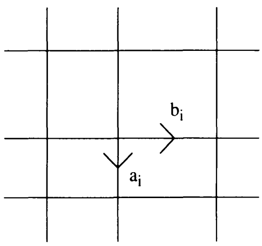
$$H^1X \times H^1X \to \mathbb{R}, \qquad [\omega] \times [\mu] \mapsto (\omega, \mu) = \iint_X \omega \wedge \mu .$$

The point is that, since the integral vanishes when either is exact, the result is independent of choice of a closed 1-form in an equivalence class. This is also a bilinear and skew-symmetric pairing.

Lemma 18.6. *Let α_i and β_i be the 1-forms constructed in* §18a. *Then $(\alpha_i, \alpha_j) = 0 = (\beta_i, \beta_j)$ for all i and j, and*

$$(\alpha_i, \beta_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Proof. As in Lemma 18.1, this is a direct calculation. The pairings that are asserted to be 0 are evident, since in these cases the wedge products of the forms are zero. This uses the fact that a wedge product $\omega \wedge \mu$ vanishes where either ω or μ vanishes, and the fact that $\omega \wedge \omega = 0$ for any 1-form ω. It therefore suffices to verify that $(\alpha_i, \beta_i) = 1$, i.e., that $\iint_X \alpha_i \wedge \beta_i = 1$, for all i. The form $\alpha_i \wedge \beta_i$ vanishes off the region that is the intersection of the two annuli around the loops a_i and b_i, and this region is diffeomorphic to a rectangle. In a coordinate patch we have the picture



On this rectangle, $\alpha_i = df_i$ and $\beta_i = dg_i$, where f_i is a function that is 0 on the right side of the rectangle and 1 on the left side, and g_i is a function that is 0 on the top of the rectangle and 1 on the bottom. If R is this rectangle,

$$\iint_X \alpha_i \wedge \beta_i = \iint_R df_i \wedge dg_i = \iint_R d(f_i \cdot dg_i) = \int_{\partial R} f_i \cdot dg_i,$$

the last step by Green's theorem on the rectangle R. Now since dg_i vanishes on the bottom and top edges of the rectangle, $f_i \cdot dg_i$ is zero on all sides of the rectangle except for the left vertical edge γ_4, where it is equal to dg_i. So

$$\int_{\partial R} f_i \cdot dg_i = -\int_{\gamma_4} dg_i = -(g_i(\gamma_4(1)) - g_i(\gamma_4(0))) = -(0-1) = 1,$$

which finishes the proof. $\square$

Exercise 18.7. For any closed 1-form ω on X, show that

$$(\alpha_j, \omega) = \int_{a_j} \omega \quad \text{and} \quad (\beta_j, \omega) = \int_{b_j} \omega.$$

Exercise 18.8. Show that, for any closed 1-forms μ and ν,

$$(\mu, \nu) = \sum_{j=1}^{g} \left(\int_{a_j} \mu \int_{b_j} \nu - \int_{a_j} \nu \int_{b_j} \mu \right).$$

Proposition 18.9. *The pairing $H^1X \times H^1X \to \mathbb{R}$ is a perfect pairing, i.e., for any linear map φ: $H^1X \to \mathbb{R}$, there is a unique ω in H^1X such that $\varphi(\mu) = (\omega, \mu)$ for all μ in H^1X.*

Taking $\varphi = 0$, this says in particular that if $\omega \in H^1X$ and $(\omega, \mu) = 0$ for all μ in H^1X, then $\omega = 0$.

Proof. In this proof we identify closed 1-forms with the classes they define in H^1X. If $\omega = \sum_{i=1}^{g}(r_i\alpha_i + s_i\beta_i)$ in H^1X, then by Lemma 18.6 we have

$$(\omega, \beta_j) = \sum_{i=1}^{g} (r_i(\alpha_i, \beta_j) + s_i(\beta_i, \beta_j)) = r_j,$$

$$(\omega, \alpha_j) = -(\alpha_j, \omega) = -\sum_{i=1}^{g} (r_i(\alpha_j, \alpha_i) + s_i(\alpha_j, \beta_i)) = -s_j,$$

The homomorphism determined by ω therefore takes α_j to $-s_j$ and β_j to r_j. If this homomorphism is zero, all s_j and r_j must be zero, which means that ω is zero in H^1X. Conversely, any homomorphism φ is determined by the values it takes on a basis of H^1X, and if we define r_j to be $\varphi(\beta_j)$ and $s_j = -\varphi(\alpha_j)$, the same equations show that $\omega = \Sigma_{i=1}^{g}(r_i\alpha_i + s_i\beta_i)$ is a class in H^1X with $(\omega, \mu) = \varphi(\mu)$ for all μ. □

Corollary 18.10. *For any γ in H_1X, there is a unique class ω_γ in H^1X such that $\int_\gamma \mu = \iint_X \omega_\gamma \wedge \mu$ for all μ in H^1X.*

Proof. Given γ, the map $\mu \mapsto \int_\gamma \mu$ is a homomorphism from H^1X to $\mathbb{R}$, so the proposition gives a unique class ω_γ with the required property. □

Exercise 18.11. (a) For $\gamma = a_i$, the corresponding class ω_{a_i} is represented by the form α_i; and for $\gamma = b_i$, the corresponding class ω_{b_i} is represented by the form β_i. This shows again that the classes of the forms α_i and β_i depend only on the choice of $a_1, \ldots, b_g$. (b) Show that the map from H_1X to H^1X taking γ to ω_γ is a homomorphism.

Problem 18.12. Suppose γ is represented by a map from the circle S^1 to X that extends to a diffeomorphism from an annulus containing S^1 to an open subset of X. Use this annulus as in §17a to construct a differential form ω that vanishes outside the image of the annulus, and, on the annulus, is the differential of a function that increases from 0 to 1 as one moves from the inside to the outside of the annulus. Show that ω represents the class ω_γ of the corollary.

We can use the corollary to define an *intersection number* $\langle\sigma, \tau\rangle$ for any σ and τ in H_1X, as we indicated at the end of the last chapter. Namely, let ω_σ and ω_τ be the cohomology classes given by the corollary, and define $\langle\sigma, \tau\rangle$ to be $(\omega_\sigma, \omega_\tau)$, i.e.,

$$\langle\sigma, \tau\rangle = \iint_X \omega_\sigma \wedge \omega_\tau = \int_\sigma \omega_\tau = -\iint_X \omega_\tau \wedge \omega_\sigma = -\int_\tau \omega_\sigma .$$

Proposition 18.13. (1) *This pairing is a bilinear skew-symmetric pairing on H_1X, i.e.,*

$$\begin{aligned} \langle m_1\sigma_1 + m_2\sigma_2, \tau\rangle &= m_1\langle\sigma_1, \tau\rangle + m_2\langle\sigma_2, \tau\rangle, \\ \langle\sigma, \mathrm{m}_1\tau_1 + m_2\tau_2\rangle &= m_1\langle\sigma, \tau_1\rangle + m_2\langle\sigma, \tau_2\rangle, \\ \langle\tau, \sigma\rangle &= -\langle\sigma, \tau\rangle, \end{aligned}$$

for all σ, τ, σ_1, σ_2, τ_1, τ_2 in H_1X and all integers m_1 and m_2.

(2) *With the basis* a_i *and* b_i *for* H_1X *as before,* $\langle a_i, a_j\rangle = \langle b_i, b_j\rangle = 0$ *for all* i *and* j, *and*

$$\langle a_i, b_j\rangle = \begin{cases} 1 & \text{if } i=j, \\ 0 & \text{if } i\neq j. \end{cases}$$

(3) *The number* $\langle \sigma, \tau\rangle$ *is always an integer.*

(4) *For any* σ *in* H_1X, *the map* $\tau \to \langle \sigma, \tau\rangle$ *is a homomorphism from* H_1X *to* $\mathbb{Z}$, *and every homomorphism from* H_1X *to* $\mathbb{Z}$ *arises in this way from a unique* σ *in* H_1X.

Proof. Part (1) follows from the corresponding assertions for the pairing (,) on cohomology classes. Part (2) follows from Lemma 18.6 and Exercise 18.11: for example, $\langle a_i, b_j\rangle = (\alpha_i, \beta_j)$, which is 1 if $i=j$ and 0 otherwise. Since the a_i and b_i form a basis for H_1X, (3) follows from (1) and (2). The proof of (4) is the same as before: the element $\sigma = \Sigma r_i a_i + s_i b_i$ has $\langle \sigma, a_j\rangle = -s_j$ and $\langle \sigma, b_j\rangle = r_j$, with this time r_i and s_i integers, and a homomorphism is determined by arbitrarily specifying the integers to which each element of a basis maps. □

Problem 18.14. Suppose σ and τ are represented by maps from S^1 to X that extend to diffeomorphisms of annuli with open sets in X, so that the images cross transversally at a finite number of points. Show how to assign numbers $+1$ or -1 to each intersection point (see the end of §17), in such a way that $\langle \sigma, \tau\rangle$ is the sum of these numbers.

It should be pointed out that, as we have constructed it here, the intersection pairing uses a differentiable structure on the surface. However, as the preceding problem indicates; it really depends only on the topology.

Problem 18.15. Prove this assertion.

18d. De Rham Theory on Surfaces

We have concentrated on the first De Rham group H^1X. As in the plane, on any surface one has the 0th group H^0X, which is the space of locally constant functions on X. If X is connected, such functions are constant, and H^0X is just the space $\mathbb{R}$ of constant functions.

We can define the *second De Rham group* H^2X of a differentiable surface X by

$$H^2X = \{\text{2-forms on } X\}/\{\text{exact 2-forms on } X\}.$$

If X is a compact oriented surface, it follows from Proposition 18.5 that integration defines a canonical map

$$H^2X \to \mathbb{R}, \qquad [\nu] \mapsto \iint_X \nu .$$

It is easy to see that this map is surjective, by constructing a form ν that vanishes outside one coordinate neighborhood $\varphi_\alpha(U_\alpha)$, and in U_α has an expression $\nu_\alpha\, dx\, dy$, with ν_α a nonnegative function with given integral. We claim that, in fact, this canonical map is an isomorphism. This is equivalent to the

Claim 18.16. *If $\iint_X \nu = 0$, then ν must be exact.*

Equivalently, *the dimension of H^2X is* (at most) *one*. With this, we will know all the De Rham cohomology groups of a compact surface of genus g:

$$H^0X = \mathbb{R}, \qquad H^1X \cong \mathbb{R}^{\oplus 2g}, \qquad H^2X \cong \mathbb{R}.$$

We will show how to extend the Mayer–Vietoris story to include the second cohomology group, which will in particular prove this claim. We will also see that if X is not compact, or if X is not orientable, then $H^2X = 0$: every closed 2-form is exact.

Exercise 18.17. Show directly from the definition that, if U is an open rectangle, every 2-form is exact. Conclude that $H^2U = 0$ if U is diffeomorphic to an open rectangle.

Exercise 18.18. Suppose a surface X is a disjoint union of a finite or infinite number of open sets U_i. Show that specifying a k-form on X is the same as specifying a k-form on each U_i, and specifying a class in H^kX is the same as specifying a class in H^kU_i for each i. In other words, H^kX is the direct product of the groups H^kU_i.

We want to compare the cohomology groups for different open sets. Note first that if U_1 is a subset of U_2, any differential function or form on U_2 determines by restriction a differential function or form on U_1. This restriction commutes with the boundary maps d (which amounts to the obvious fact that partial derivatives of a function restricted from a larger open set to a subset are the restrictions of the partial derivatives). This means that restriction takes closed forms to closed forms, and exact forms to exact forms, and hence determines linear maps, also called *restriction* maps:

$$H^k(U_2) \to H^k(U_1), \qquad k = 0, 1, 2.$$

Exercise 18.19. If $U_1 \subset U_2 \subset U_3$, show that the restriction map from $H^k(U_3)$ to $H^k(U_1)$ is the composite of the restriction map from $H^k(U_3)$ to $H^k(U_2)$ followed by the restriction map from $H^k(U_2)$ to $H^k(U_1)$.

Exercise 18.20. If U and V are open sets in a surface X, and ω is a k-form on $U \cap V$ (with $k = 1$ or 2), show that there are k-forms ω_1 on U and ω_2 on V so that $\omega = \omega_1 - \omega_2$ on $U \cap V$.

Given open sets U and V in a surface X, the contructions of Chapter 10 extend without change, giving a coboundary map

$$\delta\colon H^0(U \cap V) \to H^1(U \cup V).$$

The Mayer–Vietoris properties MV(i)–MV(v) of §10d continue to hold for these spaces and maps. Similarly, we construct a coboundary map

$$\delta\colon H^1(U \cap V) \to H^2(U \cup V).$$

The construction of a map from {closed 1-forms on $U \cap V$} to $H^2(U \cup V)$ proceeds exactly as before: Given a closed 1-form ω on $U \cap V$, use Exercise 18.20 to write ω as $\omega_1 - \omega_2$, with ω_1 and ω_2 1-forms on U and V, respectively. The 2-forms $d\omega_1$ and $d\omega_2$ agree on $U \cap V$, so there is a unique 2-form μ on $U \cup V$ that agrees with $d\omega_1$ on U and with $d\omega_2$ on V. Define $\delta(\omega)$ to be the class of this 2-form μ: $\delta(\omega) = [\mu]$. Exactly as before, one checks that this is independent of the choice of ω_1 and ω_2, and is a linear map. To see that it defines a homomorphism on the quotient space

$$H^1(U \cap V) = \{\text{closed 1-forms on } U \cap V\}/\{\text{exact 1-forms on } U \cap V\},$$

we must show that the map just defined vanishes on the exact 1-forms. If $\omega = df$, write $f = f_1 - f_2$, with f_1 and f_2 functions on U and V, respectively. Then in our construction of the coboundary of ω we may take $\omega_1 = df_1$ and $\omega_2 = df_2$, from which it follows that $d\omega_1$ and $d\omega_2$ are both zero, so $\mu = 0$ and $\delta([\omega]) = 0$, as required.

We claim next that properties (i)–(v) continue:

MV(vi). *Given ω in $H^1(U \cap V)$, $\delta(\omega) = 0$ if and only if $\omega = \alpha - \beta$ for some α in H^1U and β in H^1V.*

MV(vii). *Given μ in $H^2(U \cup V)$, μ restricts to zero in H^2U and H^2V if and only if $\mu = \delta(\omega)$ for some ω in $H^1(U \cap V)$.*

MV(viii). *Given α in H^2U and β in H^2V, α and β have the same restriction to $H^2(U \cap V)$ if and only if α and β are the restrictions of some element in $H^2(U \cup V)$.*

The proofs are exactly the same as before, and are left as exercises. In addition, we have, as an immediate consequence of Exercise 18.20:

MV(ix). *Any class μ in $H^2(U \cap V)$ can be written as the difference of a class α in H^2U and a class β in H^2V.*

Putting this together, using the fancy language as before, we have the full

Theorem 18.21 (Mayer–Vietoris Theorem). *For any open sets U and V in a surface, there is an exact sequence*

$$\begin{aligned} 0 &\longrightarrow H^0(U \cup V) \xrightarrow{+} H^0U \oplus H^0V \xrightarrow{-} H^0(U \cap V) \\ &\xrightarrow{\delta} H^1(U \cup V) \xrightarrow{+} H^1U \oplus H^1V \xrightarrow{-} H^1(U \cap V) \\ &\xrightarrow{\delta} H^2(U \cup V) \xrightarrow{+} H^2U \oplus H^2V \xrightarrow{-} H^2(U \cap V) \longrightarrow 0. \end{aligned}$$

Problem 18.22. Use Mayer–Vietoris to show that $H^2X = 0$ for any open subset of the plane.

Proposition 18.23. *If X is a surface diffeomorphic to a sphere with g handles, then dim $H^2X = 1$.*

Proof. We can write X as the union of the same two open sets U and V that we used in the proof of Proposition 17.6. Since V is diffeomorphic to an open disk, $H^1V = 0 = H^2V$, see Exercise 18.17. We leave it as an exercise to show similarly that $H^2U = 0$. We know that $H^1(U \cap V) \cong \mathbb{R}$. The proof of Proposition 17.6 shows that the map from H_1U to H_1X induced by inclusion is an isomorphism. It follows from this and Proposition 18.2 that the restriction map from H^1X to H^1U is an isomorphism. The relevant part of the Mayer–Vietoris sequence is therefore

$$H^1X \xrightarrow{\cong} H^1U \oplus 0 \to H^1(U \cap V) \to H^2X \to 0 \oplus 0.$$

It follows that the map from $H^1(U \cap V)$ to H^2X is an isomorphism. Since $H^1(U \cap V) \cong \mathbb{R}$, this shows that H^2X is one dimensional. □

Exercise 18.24. Give another proof of the proposition by writing X as a union of a sphere with $2g$ disjoint disks removed, and the union of g handles, each diffeomorphic to a cylinder $S^1 \times (0, 1)$.

Exercise 18.25. If X is a nonorientable compact surface, use Mayer–Vietoris to show that $H^2X = 0$. If X is written in normal form as in Problem 17.5, show that the dimension of H^1X is $h - 1$. Show that for all the compact surfaces, if triangulated with v vertices, e edges, and f faces,

$$v - e + f = \dim(H^0X) - \dim(H^1X) + \dim(H^2X).$$

Problem 18.26. (a) If $f\colon Y \to X$ is a differentiable map of surfaces, show how to define pull-backs $f^*\omega$ of forms from X to Y, determining maps $f^*\colon H^kX \to H^kY$ for $k = 0, 1, 2$. (b) If f is an n-sheeted covering map, show how to define push-forwards $f_*\omega$ of forms from Y to X, defining $f_*\colon H^kY \to H^kX$. (c) With f as in (b), show that the composite $f_* \circ f^*\colon H^kX \to H^kX$ is multiplication by n. (d) If X is a compact nonorientable surface, and $f\colon \tilde{X} \to X$ is its orientation covering, show that $\iint_{\tilde{X}} f^*\omega = 0$ for any 2-form ω on X. Conclude again that $H^2X = 0$.

We will see in Chapter 24 that there are corresponding homology groups H_2X, with a corresponding Mayer–Vietoris sequence, and isomorphisms $H^2X \cong \mathrm{Hom}(H_2X, \mathbb{R})$. We will also see in Chapter 24 that H^2X vanishes for any noncompact surface.

PART X

RIEMANN SURFACES

Many of the ideas about covering spaces, homology, and cohomology, can be used in the study of Riemann surfaces. A Riemann surface is a differentiable surface with a complex analytic structure. Compact Riemann surfaces arise by taking a finite-sheeted covering of the complement of a finite set in the two-sphere S^2, and then filling in appropriately over the branch points. We prove the Riemann–Hurwitz formula that computes the genus (number of handles) of a surface arising this way.

If $F(z, w)$ is an irreducible polynomial in two complex variables, it defines w as an algebraic function of z, and one is interested in the behavior of integrals of the form $\int dz/w$, or more generally $\int R(z, w)\, dz$, where R is any rational function. Associated with F there is a complex plane curve $F(z, w) = 0$; as a subset of $\mathbb{C}^2$ it is, except for a possible finite number of singularities, a Riemann surface, and the above process compactifies this surface. One of the key discoveries of the nineteenth century was how the topology of this surface, i.e., the genus g, controls much of the analysis related to the algebraic function w and its integrals. The integrals between two points are defined up to "periods," which are integrals along classes in the first homology group of the Riemann surface.

In Chapter 21 we use what we have learned about the topology and differential forms on a surface to prove the celebrated Riemann–Roch theorem, which shows how that topology (genus) of a surface controls the kinds of meromorphic functions and forms one can find on the surface. We also prove the Abel–Jacobi theorem, which similarly relates the possible integrals on the surface to its topology. We prove these theorems only for Riemann surfaces that are known to arise from a plane curve, i.e., from an irreducible polynomial $F(Z, W)$. (It is true that all compact Riemann surfaces arise this way, but we do not prove this here.) In this case there is a short proof of Weil that uses the algebraic notion of adeles, together with the few facts from analysis and topology that we have available.

CHAPTER 19

Riemann Surfaces

19a. Riemann Surfaces and Analytic Mappings

A *Riemann surface* X is a connected surface with a special collection of coordinate charts $\varphi_\alpha: U_\alpha \to X$. As before, U_α is a subset of $\mathbb{R}^2$, but now we identify $\mathbb{R}^2$ with the complex numbers $\mathbb{C}$. The requirement to be a Riemann surface is that the change of coordinate mappings $\varphi_{\beta\alpha}$ from $U_{\alpha\beta} \subset U_\alpha$ to $U_{\beta\alpha} \subset U_\beta$ are not just $\mathscr{C}^\infty$, but they must also be *analytic*, or *holomorphic*. Recall (see §9d) that an analytic function f on an open set in $\mathbb{C}$ is a complex-valued function that is locally expandable in a power series, i.e., at each point z_0 in the open set, there is a power series $\Sigma_{n=0}^{\infty} a_n(z - z_0)^n$ that converges to $f(z)$ for all z in some neighborhood of z_0. As before, another atlas of charts is compatible with a given one (and defines the same Riemann surface) if the changes of coordinates from charts in one to charts in the other are all analytic.

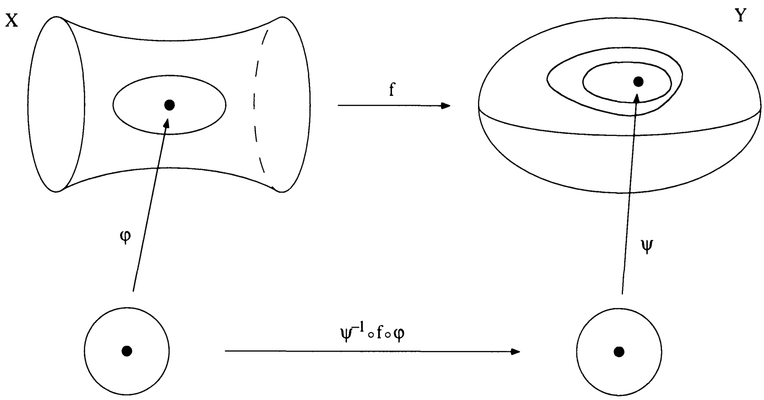
Any Riemann surface has a natural orientation. This follows from the fact that, when the analytic changes of coordinates are written out in terms of their real and imaginary parts, the Jacobian of the result has positive determinant. We will consider only connected Riemann surfaces.

Exercise 19.1. Show that in fact, if $w = f(z)$, with $w = u + iv$ and $z = x + iy$, the determinant of the Jacobian of the map from (x, y) to (u, v) is the square of the absolute value of the complex derivative $f'(z)$.

We have seen several examples of Riemann surfaces. Of course, $\mathbb{C}$ itself is one, with the identity coordinate chart $\mathbb{C} \to \mathbb{C}$, and any open subset of $\mathbb{C}$ (or any Riemann surface) is also a Riemann surface. The sphere S^2 is a compact Riemann surface, with the charts given by spherical projection from the north and south poles. Indeed, we saw (Exercise 7.14) that the change of coordinates mapping in this case is given by the map $z \mapsto 1/z$ from $\mathbb{C} \setminus \{0\}$ to $\mathbb{C} \setminus \{0\}$. (When we speak of S^2, it will always be regarded as a Riemann surface with these coordinate charts.) The torus $\mathbb{R}^2/\mathbb{Z}^2 = \mathbb{C}/(\mathbb{Z} + \mathbb{Z}i)$ is a compact Riemann surface, using the projections from small open sets in $\mathbb{C}$ to the quotient for coordinate charts.

Any compact oriented surface can be given a structure of a Riemann surface (as we shall see later), but, except for the sphere S^2, there are infinitely many nonequivalent Riemann surfaces with the same underlying surface. For example, if $\Lambda \subset \mathbb{C}$ is any lattice, i.e., a subgroup generated by two elements that give a basis for $\mathbb{C}$ as a real vector space, then $\mathbb{C}/\Lambda$ is likewise a compact Riemann surface. These are all homeomorphic (and diffeomorphic) to each other, but they are generally not isomorphic Riemann surfaces (see Exercise 19.17). This is part of the general subject of "moduli of Riemann surfaces"—the study of the set of all Riemann surfaces of a given genus—that we can only hint at in this book.

If X is a Riemann surface, it makes sense to say that a function $f: X \to \mathbb{C}$ is *analytic* (or *holomorphic*): f is analytic if for each coordinate chart $\varphi_\alpha: U_\alpha \to X$, the composite $f \circ \varphi_\alpha$ is an analytic function on the open set U_α in $\mathbb{C}$. More generally, if X and Y are Riemann surfaces, a mapping $f: X \to Y$ is *analytic* at a point P in X if there are charts $\varphi: U \to X$ and $\psi: V \to Y$ mapping to neighborhoods of P and $f(P)$, respectively, so that $f(\varphi(U)) \subset \psi(V)$, and the composite $\psi^{-1} \circ f \circ \varphi$ is an analytic function from U to V:



This condition is independent of the choice of coordinates φ and ψ. In fact, we can choose both U and V to be disks centered at the origins, with these origins mapped to P and $f(P)$. In this case the composite $h = \psi^{-1} \circ f \circ \varphi$ has the form $h(z) = \Sigma_{n=1}^{\infty} a_n z^n$, for some converging power series. As we will soon see, the order of vanishing of h at the origin, i.e., the smallest integer e such that $a_e \neq 0$, is independent of the choice of coordinates. This integer is called the *ramification index* of f at P, and we will denote it by $e_f(P)$, or just $e(P)$ when one function f is being considered. (If f is constant, of course, h is identically 0, and $e = \infty$, but we will not be interested in constant maps.) The point P is called a *ramification point* for f if $e_f(P) > 1$.

We claim next that we can change the coordinate chart φ so that the composite $\psi^{-1} \circ f \circ \varphi$ is the function $z \mapsto z^e$. To see this, write $h(z) = z^e \cdot g(z)$, where $g(z)$ is an analytic function at the origin and $g(0) \neq 0$. ($g(z) = \Sigma_{n=0}^{\infty} a_{n+e} z^n$.) Such an analytic function g, possibly in a smaller disk around the origin, can be written as the eth power of an analytic function $k(z)$; for example, $g(z)/a_e$ maps 0 to 1, so maps a neighborhood of 0 to the right half-plane, so one can compose it with a branch of the log function

$$\log(z) \;=\; -\sum_{n=1}^{\infty} \frac{1}{n}(1-z)^n \qquad \text{for} \quad |z-1| < 1,$$

and then set

$$k(z) \;=\; \alpha \cdot \exp\left(\frac{1}{e}\log(g(z)/a_e)\right),$$

where α is any eth root of a_e and $\exp(z) = \Sigma_{n=0}^{\infty}(1/n!)z^n$.

Exercise 19.2. Verify that $k(z)^e = g(z)$ in some disk containing the origin. Show that there are exactly e choices for such k (up to shrinking the disks they are defined on), obtained by the e choices for the eth root α of a_e.

Now we have written $h(z) = (z \cdot k(z))^e$, where k is an analytic function with $k(0) \neq 0$. The mapping $z \mapsto z \cdot k(z)$ is an analytic isomorphism in some neighborhood of the origin, since its derivative does not vanish at the origin. Therefore, we can define a new coordinate chart $\tilde{\varphi}$ (from some small disk U' to X) so that $\tilde{\varphi}(z) = \varphi(z \cdot k(z))$ for all small z. It follows that, with this new coordinate chart $\tilde{\varphi}$, the composite $\psi \circ f \circ \tilde{\varphi}$ is just the map $z \mapsto z^e$, as required.

The mapping $z \mapsto z^e$ is familiar: it maps 0 to 0, and outside the origin it is an e-sheeted covering map. This shows that there are neighborhoods U of P and V of $f(P)$ so that f maps U to V, and the mapping $U \setminus \{P\} \to V \setminus \{f(P)\}$ is an e-sheeted covering. This shows in particular that the number e is independent of choices, and in fact depends only on the topology of the map f near P.

Note another important consequence of this local structure: a nonconstant analytic mapping is always an *open* mapping, i.e., the image of any open set is open. In particular, if X is compact and Y is not compact, the only analytic mappings from X to Y are constants. When $Y = \mathbb{C}$ this openness is a strong form of the maximum principal: one cannot have a point P_0 such that $|f(P_0)| \geq |f(P)|$ for P in a neighborhood of P_0.

Proposition 19.3. *Let $f\colon X \to Y$ be a nonconstant analytic map between compact Riemann surfaces.*

(1) *There are a finite number of ramification points. Let $R \subset X$ be the set of ramification points, and set $S = f(R) \subset Y$.*

(2) *The map from $X \setminus f^{-1}(S)$ to $Y \setminus S$ determined by f is an n-sheeted covering map for some finite number n. This integer n is called the degree of f.*

(3) *For any point Q in Y, $\Sigma_{P \in f^{-1}(Q)} e_f(P) = n$.*

Proof. (1) follows from the fact that for any point P, there is a neighborhood of P that contains no other ramification point, using the compactness of X to cover it by a finite number of such neighborhoods. It follows similarly that $f^{-1}(Q)$ is finite for all points Q in Y (consider a possible limit point of the set $f^{-1}(Q)$).

To prove (2), let $Q \notin S$, and $f^{-1}(Q) = \{P_1, \ldots, P_n\}$. There are neighborhoods U_i of P_i and V_i of Q such that f maps U_i homeomorphically onto V_i. Shrinking the U_i if necessary, we may assume they are disjoint, and that V_i contains no point of S. For any connected neighborhood V of Q contained in $V_1 \cap \ldots \cap V_n$, let $U_i' = U_i \cap f^{-1}(V)$; note that f maps each U_i' homeomorphically onto V. We claim that for V sufficiently small, $f^{-1}(V)$ is the disjoint union of the sets U_i', from which it follows that V is evenly covered, so f is a covering in a neighborhood of Q. To prove this claim, suppose on the contrary that there is a sequence of neighborhoods N_i of Q whose intersection is $\{Q\}$, such that there is a point P_i' in $f^{-1}(N_i)$ with P_i' not in $U_1 \cup \ldots \cup U_n$. By the compactness of X, a subsequence of these P_i' must have a limit point P' in X. By the continuity of f, $f(P') = Q$, so $P' = P_j$ for some j. But this contradicts the assumption that the points

P_i' are not in U_j for all i and j. (In fact, having proved (2), it follows that the whole neighborhood $V = \cap V_i$ is evenly covered.)

For (3), again let $f^{-1}(Q) = \{P_1, \ldots, P_m\}$, and find neighborhoods U_i of P_i and V_i of Q such that f maps U_i onto V_i, so that in local coordinates it is the map $z \mapsto z^{e(P_i)}$, so it is $e_f(P_i)$ to 1 except at the point P_i. Again if V is a neighborhood of Q contained in the intersection of the V_i, but not containing any other point of S, then there are $\Sigma e_f(P_i)$ points over a point Q' in V except for the point Q. By (2), this sum must be the number n of sheets of the covering. □

A *meromorphic function* f on a Riemann surface X is the same as an analytic function $f: X \to S^2$ from X to the Riemann sphere. Equivalently, f is an analytic function from $X \setminus S$ to $\mathbb{C}$, with S a discrete subset of X, and for each point P in X, there is a coordinate chart $\varphi: U \to X$ taking 0 to P, so that on $U \setminus \{P\}$, $f \circ \varphi(z) = z^k \cdot g(z)$, for some integer k and some analytic function g that is nonvanishing in U. The integer k is called the *order* of f at P, and is denoted $\mathrm{ord}_P(f)$. It is independent of the choice of the local coordinate. If k is positive, one says that f has a *zero* of order k, or *vanishes to order* k, and if k is negative, we say that f has a *pole of order* $-k$.

Exercise 19.4. Show that f is meromorphic at P if and only if there is a coordinate chart as above and an integer k so that the function $z^{-k} \cdot (f \circ \varphi)(z)$ is bounded as z approaches 0.

In terms of the mapping $f: X \to S^2$, the order of f at P is positive if $f(P) = 0$, negative if $f(P) = \infty$, and zero otherwise. If $f(P) = 0$, the order of f at P is just the ramification index $e_f(P)$ of f at P. Similarly, if $f(P) = \infty$, the order of f at P is minus the ramification index at P, since $1/z$ is a local parameter for the Riemann sphere at ∞. If X is compact, we know that the sum of the ramification indices over any point in S^2 is the degree n of f. Assertion (3) of the proposition says that f takes on all values the same number of times, counting multiplicity correctly. In particular, for the values 0 and ∞, this gives:

Corollary 19.5. *For any nonconstant meromorphic function f on a compact Riemann surface X,*

$$\sum_{P \in X} \mathrm{ord}_P(f) \;=\; 0.$$

Exercise 19.6. Let $p_1(z)$ and $p_2(z)$ be polynomials in z of degrees d_1 and d_2, with no common factors. (a) Show that $f(z) = p_1(z)/p_2(z)$ determines an analytic mapping from $S^2 = \mathbb{C} \cap \{\infty\}$ to itself. (b) Show that the degree of this mapping is the maximum of the integers d_1 and d_2. (c) Show that the ramification index of this mapping at ∞ is $|d_1 - d_2|$. (d) Show that every analytic mapping from S^2 to S^2 has this form.

19b. Branched Coverings

If Y is a Riemann surface, and $p: X \to Y$ is a covering mapping, then there is a unique structure of a Riemann surface on X so that p is an analytic mapping. In fact, one can choose charts $\varphi_\alpha: U_\alpha \to Y$ for Y so that each $\varphi_\alpha(U_\alpha)$ is evenly covered by p; each component $V_{\alpha,i}$ of $p^{-1}(\varphi_\alpha(U_\alpha))$ maps homeomorphically to $\varphi_\alpha(U_\alpha)$ by p, and the composite

$$U_\alpha \to \varphi_\alpha(U_\alpha) \to V_{\alpha,i}$$

of φ_α and the inverse of p, gives a coordinate chart on X. It is straightforward to verify that the changes of coordinates for these charts are analytic, so define a Riemann surface structure on X.

Exercise 19.7. If Y is a Riemann surface, and G is a group acting evenly on Y by analytic isomorphisms, show that Y/G can be given the structure of a Riemann surface so that $Y \to Y/G$ is analytic.

An important case of this is the fact that the universal covering $\widetilde{X}$ of a Riemann surface is a Riemann surface, and the fundamental group $\pi_1(X, x)$ acts as a group of analytic isomorphisms of $\widetilde{X}$. When $X = S^2$, $\widetilde{X} = X$; when $X = \mathbb{C}/\Lambda$ for a lattice Λ, $\widetilde{X} = \mathbb{C}$. It is a fact of complex analysis (the *uniformization theorem*) that in all other cases, the universal covering is isomorphic to the upper half plane H (or an open disk). The automorphisms of H all have the form $z \mapsto (az+b)/(cz+d)$ where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is a real matrix of determinant 1. This means that X can be realized as the quotient of H by a subgroup of $\mathrm{SL}_2(\mathbb{R})/\{\pm I\}$ acting evenly on H. We will not have more to say about this situation, which is a fundamental area of mathematics in its own right.

In the last section we saw that any (nonconstant) analytic map $X \to Y$ of compact Riemann surfaces determines a finite-sheeted covering $X \setminus f^{-1}(S) \to Y \setminus S$, for some finite set S in Y. Our goal in this section is to reverse this process. Given a Riemann surface Y, a finite subset S of Y, and a finite-sheeted topological covering $p: X^\circ \to Y \setminus S$, by what we just proved, X° gets a structure of a Riemann surface so that this mapping is analytic. We want to "fill in" the missing points over the points of S, embedding X° in a Riemann surface X, so that we have an analytic mapping f from X to Y compatible with the given covering:

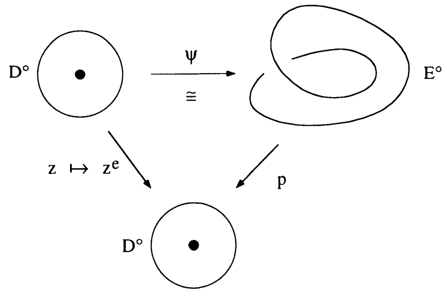
$$\begin{tikzcd} X^\circ \arrow[r, phantom, "\subset"] \arrow[d, "p"'] & X \arrow[d, "f"] \\ Y \setminus S \arrow[r, phantom, "\subset"] & Y. \end{tikzcd}$$

If Y is compact, we want X to be compact. In general, we want the mapping f to be *proper*: for any compact set K of Y, the inverse image $f^{-1}(K)$ should be a compact set in X. The problem is local on Y: we need to fill in the covering over each point of S. We will look first at the "local model" of a Riemann surface—an open disk.

Let $D = \{z\colon |z| < 1\}$, $D^\circ = D \setminus \{0\}$. We know all about the coverings of D°. In fact, its fundamental group is $\mathbb{Z}$ (since it contains a circle as a deformation retract), so connected finite-sheeted coverings correspond to subgroups of finite index in $\mathbb{Z}$, and the only such subgroups of $\mathbb{Z}$ are the groups $e\mathbb{Z}$ for e a positive integer. The e-sheeted covering corresponding to this subgroup is

$$p_e\colon D^\circ \to D^\circ, \qquad z \mapsto z^e.$$

This means that if $p\colon E^\circ \to D^\circ$ is any e-sheeted connected covering, there is a homeomorphism $\psi\colon D^\circ \to E^\circ$ so that $p \circ \psi = p_e$.

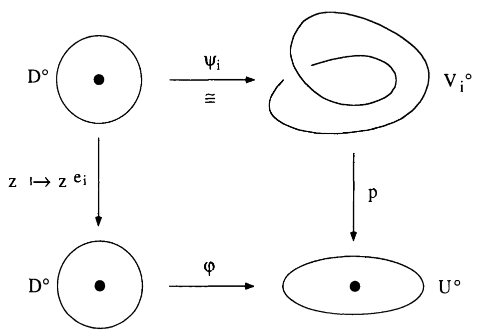


This homeomorphism ψ is not uniquely determined, depending, as we know, on a choice of where a base point is mapped. In particular, there are exactly e such homeomorphisms, corresponding to the e points of E° over a given base point. The other choices of ψ have the form $z \mapsto \psi(\zeta_k \cdot z)$, where $\zeta_k = \exp(2\pi k i/e)$ is one of the eth roots of unity, $1 \le k \le e-1$.

It is now clear how to fill in the covering $p\colon E^\circ \to D^\circ$. Define E to be the union of E° with one other point, and put the structure of a

Riemann surface on E so that the extension of ψ from D° to D, mapping 0 to the added point, is an isomorphism of D with E. Note that this does not depend on the choice of ψ, since the map $z \mapsto \zeta_k \cdot z$ is an analytic isomorphism.

Now return to the given covering $p\colon X^\circ \to Y \setminus S$. Given Q in S, we can find a coordinate chart $\varphi\colon D \to \varphi(D) = U \subset Y$, with D the open unit disk in the plane, and $\varphi(0) = Q$, such that U does not contain any other point of S but Q. Let $U^\circ = U \setminus \{Q\}$. The covering p restricts to a covering of U°, so $p^{-1}(U^\circ)$ is a disjoint union of connected open sets $V_1^\circ, \ldots, V_m^\circ$, with each $V_i^\circ \to U^\circ$ a connected covering, say with e_i sheets. By what we just saw, one can find homeomorphisms $\psi_i\colon D^\circ \to V_i^\circ$ such that the diagram



commutes, i.e., $p(\psi_i(z)) = \varphi(z^{e_i})$, with ψ_i unique up to first multiplying z by an e_ith root of unity. We can therefore add one point to each V_i°, getting spaces V_i so that each ψ_i extends to a homeomorphism from D to V_i. Taking these extensions as charts, the disjoint union of the V_i becomes a Riemann surface. The map from V_i° to U° extends to an analytic map from V_i to U that has ramification index e_i at the added point.

If this is done at each point of S, one gets a space X that is the union of X° with a finite number of points. These local charts give X the structure of a Riemann surface (noting that the added charts are compatible with the given charts on X°), and the covering p is extended to an analytic mapping $f\colon X \to Y$.

Exercise 19.8. Verify that this defines a Riemann surface X, and the map f is analytic and proper.

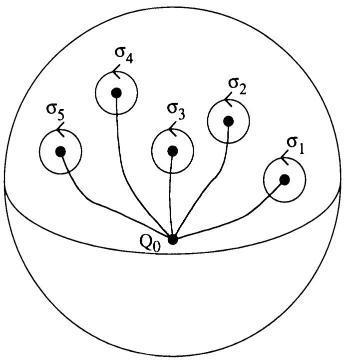
These results are summarized in the following proposition:

Proposition 19.9. *Let Y be a Riemann surface, S a finite subset of Y, and $p\colon X^\circ \to Y \setminus S$ a finite-sheeted covering map, with X° connected. Then there is an embedding of X° as an open subset of a Riemann surface X that is a union of X° and a finite set, so that p extends to a proper analytic mapping from X to Y.*

Exercise 19.10. Show that X, with its Riemann surface structure, is unique up to canonical isomorphism: if $X^\circ \subset X' \to Y$ is another, there is a unique isomorphism of Riemann surfaces from X to X' compatible with the inclusions of X° and mappings to Y.

Exercise 19.11. Extend the proposition to the case where S is an infinite but discrete subset of Y.

One reason for the importance of this proposition is that, since we know the fundamental group of $Y \setminus S$ (see Problem 17.12 for the general case), we know all possible finite-sheeted coverings. The main case of concern to us will be $Y = S^2$, with S a set of r points. By choosing disjoint arcs from a base point Q_0 to these points, we can number the points $S = \{Q_1, \ldots, Q_r\}$ so that the arcs to them occur in this order, going counterclockwise around Q_0. Then one can construct loops $\sigma_1, \ldots, \sigma_r$ that go from Q_0 along an arc to a point near Q_i, make a counterclockwise circle around Q_i, and go back along the arc to Q_0.



Exercise 19.12. Show that $\pi_1(S^2 \setminus \{Q_1, \ldots, Q_r\}, Q_0)$ is the free group F_r on the generators $\sigma_1, \ldots, \sigma_r$, modulo the least normal subgroup containing $\sigma_1 \cdot \ldots \cdot \sigma_r$. This is isomorphic to a free group on $r-1$ generators $\sigma_1, \ldots, \sigma_{r-1}$.

Remark 19.13. In §16d we described a correspondence between n-sheeted coverings of a space and actions of its fundamental group on a finite set T with n elements. To give an action of this fundamental group on a set T is the same as giving r permutations $s_1, \ldots, s_r$ of T, with the requirement that $s_1 \cdot \ldots \cdot s_r$ is the identity permutation. So any such data determine a branched covering of the sphere. The classical way to do this was to use the fact that the complement of the r arcs drawn is simply connected, so a covering over this is a disjoint union of r copies of it. Labeling these by elements of T, the permutations are used to describe how to glue the sheets together across the arcs, to get the Riemann surface X. (Note, however, that the arcs are only a tool used to give nice generators for the fundamental group, and are not necessary for describing the covering.)

Exercise 19.14. (a) Carry out this construction, and show that it agrees with that described before. (b) Show that the covering corresponding to this choice of permutations is connected if and only if they generate a transitive subgroup of the automorphisms of T. (Recall that a group of permutations is *transitive* if, for any elements t_1 and t_2 of T, there is a permutation in the group that takes t_1 to t_2.) (c) Let X be the corresponding Riemann surface, with $f\colon X \to S^2$ the analytic mapping. For each i, write T as a disjoint union of sets $T_{i,j}$ so that s_i permutes the elements of each $T_{i,j}$ cyclically. Show that there is one point of X over Q_i for each set $T_{i,j}$, and the cardinality of $T_{i,j}$ is the ramification index of f at this point.

19c. The Riemann–Hurwitz Formula

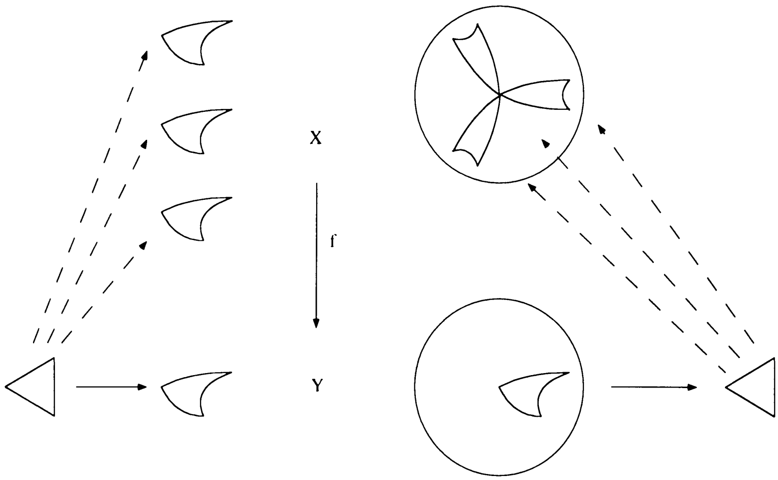
Given an analytic mapping $f\colon X \to Y$ between compact Riemann surfaces, our next aim is to describe the topology of X in terms of the topology of Y and the local behavior around the branch points.

Theorem 19.15 (Riemann–Hurwitz). *Let $f\colon X \to Y$ be an analytic map of degree n between compact Riemann surfaces. If Y can be triangulated, so can X, and the genus g_X of X and the genus g_Y of Y are related by the formula*

$$2g_X - 2 \;=\; n(2g_Y - 2) + \sum_{P \in X} (e_f(P) - 1)\,.$$

Proof. By refining the triangulation of Y if necessary, for example, using the barycentric subdivision described in §8b, we can assume the triangulation has the property that each point Q of S is a vertex in the triangulation. We will construct from this an explicit triangulation of X. It is simplest to describe the vertices and the "open" edges (homeomorphic to $(0, 1)$) and the open faces (homeomorphic to the interior of a plane triangle) of this triangulation of X: these are exactly the components of the inverse images by f of the corresponding vertices, open edges, and open faces of the triangulation of Y. The point is that each of the open edges and faces on Y is in the locus where the map is a covering, and, since edges and faces are simply connected, the restriction of the covering to each of them is trivial, which means that there are n subsets of X that map homeomorphically onto each of them, and these are the specified open edges and faces.

For each closed edge and face on Y the triangulation specifies a homeomorphism with a closed interval or a closed triangle. Composing with the projections, we have compatible homeomorphisms of the open edges and faces on X with the interiors of these intervals or triangles. To see that we have a triangulation, we must verify that these homeomorphisms extend to the closed intervals and triangles. This is clear when the image edges or triangles in Y do not have vertices in S, since in that case the coverings are trivial over the entire closed triangle. The same argument shows that the homeomorphisms extend (uniquely, by continuity) to the closures of the intervals and triangles, except possibly in an arbitrarily small neighborhood of a vertex mapping to a point in $f^{-1}(S)$. To see that it extends continuously to such points is a local question, so we can assume we are in the situation $z \mapsto z^e$ of the mapping from the disk to itself. The corresponding vertex is sent to the center of the disk, and the continuity of the map at the vertex follows from the fact that $|z| \to 0$ if and only if $|z^e| \to 0$. The following picture shows the two types of behavior:



Now suppose the triangulation of Y has v vertices, e edges, and f faces. By the construction of the triangulation of X, it has $n \cdot f$ faces and $n \cdot e$ edges. But the number of vertices is not $n \cdot v$, since over each point Q of S there may be fewer than n points. In fact, by the equation $\Sigma e_f(P) = n$, the sum taken over the points P in $f^{-1}(Q)$, the number of points over Q is $n - \Sigma(e_f(P) - 1)$. It follows that the number of vertices in the triangulation of X is $n \cdot v - \Sigma(e_f(P) - 1)$, with the sum over all P in $f^{-1}(S)$, or in X, since $e_f(P) = 1$ if P is not in $f^{-1}(S)$. The Euler characteristic of X is therefore

$$n \cdot v - \sum_{P \in X} (e_f(P) - 1) - n \cdot e + n \cdot f \;=\; n(v - e + f) - \sum_{P \in X} (e_f(P) - 1).$$

Replacing the left side by $2 - 2g_X$, and $v - e + f$ by $2 - 2g_Y$, the formula of the proposition results. $\square$

Exercise 19.16. Give an alternative proof of the theorem by triangulating Y so that each point of S lies inside a face, and no face contains two points of S.

For example, when $Y = S^2$, we have Riemann's formula

$$\sum_{P \in X} (e_f(P) - 1) \;=\; 2g_X + 2n - 2.$$

If X is also the sphere, this gives

$$\sum_{P \in X} (e_f(P) - 1) \;=\; 2n - 2.$$

The Riemann–Hurwitz formula can be useful to limit the possibilities for mappings between Riemann surfaces. Consider for example a nonconstant analytic mapping $f\colon \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$, where Λ and Λ' are two lattices in $\mathbb{C}$. Since both have genus 1, it follows from the Riemann–Hurwitz formula that $e_f(P) = 1$ for all P, i.e., f must be unramified. Now f lifts to an analytic mapping $\tilde{f}\colon \mathbb{C} \to \mathbb{C}$ so that the diagram

$$\begin{tikzcd} \mathbb{C} \arrow[r, "\tilde{f}"] \arrow[d, "p"'] & \mathbb{C} \arrow[d, "p'"] \\ \mathbb{C}/\Lambda \arrow[r, "f"] & \mathbb{C}/\Lambda', \end{tikzcd}$$

commutes. In fact, since $\mathbb{C}$ is simply connected, the composite $f \circ p$ lifts through the covering p' by Proposition 13.5, to produce a continuous map $\tilde{f}$, and $\tilde{f}$ is automatically analytic since the projections from $\mathbb{C}$ to the quotient spaces are local isomorphisms. Similarly, since f is unramified, $\tilde{f}$ is also unramified. From the topology one can see also that $\tilde{f}$ extends continuously to a map from $S^2 = \mathbb{C} \cap \{\infty\}$ to itself, taking ∞ to ∞. From Exercise 19.4 this extension, also denoted $\tilde{f}$, is an analytic mapping from S^2 to S^2. If the degree of $\tilde{f}$ is n, the sum of the $e_f(P) - 1$ can be at most $n - 1$ (since ramification can take place only over ∞), and this contradicts the Riemann–Hurwitz formula unless $n = 1$. By Exercise 19.6, we deduce that $\tilde{f}(z) = \lambda z + \mu$ for some complex numbers λ and μ, with $\lambda \neq 0$. In particular, $\lambda \cdot \Lambda + \mu \subset \Lambda'$. Conversely, any such λ and μ determine an analytic mapping f. This puts strong restrictions on the relations between the lattices, and on the possible maps f.

Exercise 19.17. (a) Show that for any lattice Λ there is a nonzero complex number a so that $a \cdot \Lambda$ is generated by 1 and τ, where τ is a number in the upper half plane. So every $\mathbb{C}/\Lambda$ is isomorphic to one of the form $\mathbb{C}/(\mathbb{Z} + \mathbb{Z} \cdot \tau)$. (b) Show that for two numbers τ and τ' in the upper half plane, $\mathbb{C}/(\mathbb{Z} + \mathbb{Z} \cdot \tau)$ is isomorphic to $\mathbb{C}/(\mathbb{Z} + \mathbb{Z} \cdot \tau')$ if and only if $\tau' = (a\tau + b)/(c\tau + d)$ for some integers a, b, c, d with $ad - bc = 1$. (c) Show that the only analytic maps from the Riemann surface $\mathbb{C}/(\mathbb{Z} + \mathbb{Z} \cdot \tau)$ to itself that take the image of 0 to itself are given by multiplication by an integer n, unless τ satisfies a quadratic polynomial with integer coefficients.

It follows from the preceding exercise that the set of compact Riemann surfaces that are isomorphic to some $\mathbb{C}/\Lambda$ are in one-to-one correspondence with the space $\mathrm{H}/\mathrm{SL}_2(\mathbb{Z})$ of orbits of the discrete group

$SL_2(\mathbb{Z})$ on the upper half plane H. In particular, not all such Riemann surfaces are isomorphic as Riemann surfaces, although they are all diffeomorphic to $S^1 \times S^1$. We will see later that every compact Riemann surface of genus 1 has the form $\mathbb{C}/\Lambda$, so $H/SL_2(\mathbb{Z})$ is the moduli space of compact Riemann surfaces of genus 1.

Exercise 19.18. Let $f: X \to Y$ be a nonconstant analytic map between compact Riemann surfaces. Show that $g_X \geq g_Y$, and $g_X > g_Y$ unless $g_Y = 0$ or $g_Y = 1$ or f is an isomorphism.

CHAPTER 20

Riemann Surfaces and Algebraic Curves

20a. The Riemann Surface of an Algebraic Curve

If $F(Z, W)$ is a polynomial in two variables, with complex coefficients, that is not simply a constant, its zero set

$$C = \{(z, w) \in \mathbb{C}^2 \colon F(z, w) = 0\}$$

is called a "complex affine plane curve." Identifying $\mathbb{C}^2$ with $\mathbb{R}^4$, C is defined by two real equations: the vanishing of the real and imaginary parts of $F(z, w)$. We may therefore expect C to be a surface, and this expectation is generally true, except that, just as in the case of real curves, C may have singularities. We will use the construction of the preceding section to "remove" these singularities, and also add some points over them and "at infinity," to get a compact Riemann surface. In fact, if F is not irreducible, the surface we get will be the disjoint union of the surfaces we get from the irreducible factors of F, so we assume for now that F is an irreducible polynomial, i.e., it has no nontrivial factors but constants. Write

$$F(Z, W) = a_0(Z)W^n + a_1(Z)W^{n-1} + \ . \ . \ . + a_{n-1}(Z)W + a_n(Z),$$

with $a_i(Z)$ a polynomial in Z alone, and $a_0(Z) \neq 0$. We may also assume that n is positive, for otherwise $F = bZ + c$, and C is isomorphic to $\mathbb{C}$, given by the projection to the second factor. We will need a little piece of algebra. Let $F_w = \partial F/\partial W$, which is a polynomial of degree $n-1$ in W:

$$F_w = \frac{\partial F}{\partial W} = n \cdot a_0(Z)W^{n-1} + (n-1) \cdot a_1(Z)W^{n-2} + \ldots + a_{n-1}(Z).$$

Lemma 20.1. *There are polynomials $B(Z, W)$, $C(Z, W)$, and $d(Z)$, with $d(Z) \neq 0$, so that*

$$B(Z, W) \cdot F(Z, W) + C(Z, W) \cdot F_w(Z, W) = d(Z).$$

Proof. We use the lemma of Gauss (Appendix C3): if F is irreducible in $\mathbb{C}[Z, W]$, then F is also irreducible in $\mathbb{C}(Z)[W]$, where $\mathbb{C}(Z)$ is the field of rational functions in Z. The equation of Lemma 20.1 can be found (and computed) from the Euclidean algorithm in the ring $\mathbb{C}(Z)[W]$, as follows. Divide F by F_w to get, after clearing denominators, an equation of polynomials

$$b_0 \cdot F = Q_1 \cdot F_w + R_1,$$

where $b_0 \in \mathbb{C}[Z]$, and R_1 is a polynomial of degree less than $n-1$ in W. If the degree of R_1 in W is positive, divide F_w by R_1, getting an equation

$$b_1 \cdot F_w = Q_2 \cdot R_1 + R_2;$$

continuing, find equations $b_i \cdot R_{i-1} = Q_{i+1} \cdot R_i + R_{i+1}$, until for some k, R_{k+1} has degree 0 in W. Note that $R_{k+1} \neq 0$, since otherwise R_k would divide R_{k-1}, then $R_{k-2}, \ldots,$ and finally F_w and F, contradicting the fact that F is irreducible.

Set $d = R_{k+1}$. To find an equation of the form required amounts to showing that d is in the ideal in $\mathbb{C}[Z, W]$ generated by F and F_w; this ideal contains R_1 by the first equation, then R_2 by the second, and so on, until finally it contains $R_{k+1} = d$. □

If one takes an equation as in the lemma so that B, C, and d are not all divisible by any nonconstant polynomial in Z, then d will be unique up to multiplication by a constant; this d is called the *discriminant* of F with respect to W. For our purposes any d will do.

Exercise 20.2. For $F = W^2 + b(Z)W + c(Z)$, show that $d(Z) = b(Z)^2 - 4c(Z)$. For $F = W^3 + b(Z)W + c(Z)$, show that $d(Z) = 4b(Z)^3 + 27c(Z)^2$.

Let p be the first projection from C to $\mathbb{C}$: $p(z, w) = z$. We will show that, if a finite number of points are removed, p becomes a covering map.

Lemma 20.3. *There is a finite subset S of $\mathbb{C}$ such that the projection from $C \setminus p^{-1}(S)$ to $\mathbb{C} \setminus S$ is a finite covering with n sheets.*

Proof. From Lemma 20.1 we draw the conclusion that if $z \in \mathbb{C}$ and $d(z) \neq 0$, then there is no w with $F(z, w) = 0$ and $F_w(z, w) = 0$. This means that the equation $F(z, W) = 0$ has no multiple roots. If in addition $a_0(z) \neq 0$, then this equation has n distinct roots. We take S to be the set where $a_0(z) \cdot d(z) \neq 0$.

Suppose $z_0 \notin S$, and let $w_1, \ldots, w_n$ be the roots of the equation $F(z_0, W) = 0$. We want to find analytic functions $g_1, \ldots, g_n$ defined in a neighborhood of z_0 so that $g_i(z_0) = w_i$ and $F(z, g_i(z)) \equiv 0$. One way to do this is to apply the implicit function theorem, either for a subset of $\mathbb{R}^4$ defined by two real equations, or the complex analogue for a subset of $\mathbb{C}^2$ defined by one equation. Another is to use the Argument Principle. For this, take small disjoint closed disks around these n points, and let γ_i be a counterclockwise path around the boundary of the disk around w_i. For z near z_0 and w on a circle γ_i, $F(z, w) \neq 0$ by continuity. For z near z_0,

$$\frac{1}{2\pi i}\int_{\gamma_i} \frac{F_w(z, w)}{F(z, w)} dw = 1,$$

since the integral is an integer that is 1 when $z = z_0$, and it varies continuously with z. This means that there is exactly one root of the equation $F(z, W) = 0$ inside γ_i. In fact, from Problem 9.34, this root is given by the formula

$$g_i(z) = \frac{1}{2\pi i}\int_{\gamma_i} w \frac{F_w(z, w)}{F(z, w)} dw.$$

In particular, take U to be a disk around z_0 where all these functions g_i are defined, and set $V_i = \{(z, g_i(z)) \colon z \in U\}$. Since these points give all possible roots of $F(z, W) = 0$ over z in U, we see that $p^{-1}(U)$ is the union of open sets V_i, and p maps each V_i homeomorphically onto V, with an inverse given by $z \mapsto (z, g_i(z))$. □

Now regard $\mathbb{C} \subset S^2 = \mathbb{C} \cup \{\infty\}$ as usual, and enlarge S by including the point at infinity in it. By the lemma, we have a covering map $X^\circ \to S^2 \setminus S$, with $X^\circ = \{(z, w) \colon F(z, w) = 0 \text{ and } z \notin S\}$. This covering map gives X° the structure of a Riemann surface, except that we must prove it is connected:

Lemma 20.4. *If $F(Z, W)$ is irreducible, then X° is connected.*

Proof. Let Y° be a connected component of X°. If X° is not connected, then the covering $X^\circ \to S^2 \setminus S$ restricts to a covering $Y^\circ \to S^2 \setminus S$ with $m < n$ sheets (see Exercise 11.11). For each z not in S, let $e_1(z)$, $e_2(z), \ldots, e_m(z)$ be the elementary symmetric functions in the m values of w on the points in $p^{-1}(z) \cap Y^\circ$. That is, $e_1(z)$ is the sum of these m values, $e_2(z)$ is the sum of all products of pairs of these values, and so on until $e_m(z)$ is the product of all m values. These functions e_i are clearly analytic on $S^2 \setminus S$. They are in fact meromorphic on S^2, as follows from the fact that, for a in S, after multiplying by some $(z-a)^k$, the roots $w_i(z)$ approach 0 (see Exercise 19.4). Consider the polynomial

$$G = W^m - e_1(Z)W^{m-1} + e_2(Z)W^{m-2} - \ldots + (-1)^m e_m(Z)$$

in $\mathbb{C}(Z)[W]$. For any z not in S,

$$G(z, W) = \prod_{P \in p^{-1}(z) \cap Y^\circ} (W - w(P)),$$

$$F(z, W) = a_0(z) \prod_{P \in p^{-1}(z)} (W - w(P)).$$

It follows that G divides F in $\mathbb{C}(Z)[W]$, since the remainder obtained by dividing F by G would be a polynomial whose coefficients are rational functions of Z vanishing on an infinite set. This contradicts the irreducibility of F, and completes the proof. □

By Proposition 19.9 this covering can be extended to a compact Riemann surface X, together with an analytic map $f\colon X \to S^2$. This Riemann surface is called *the Riemann surface of the algebraic curve C* or of the polynomial F. It is easily seen to be independent of choice of S. In fact, the following exercise shows more.

Exercise 20.5. A point $P = (z_0, w_0)$ is called a *nonsingular* point of C if either: (i) $F_w(z_0, w_0) \neq 0$; or (ii) $F_z(z_0, w_0) \neq 0$; or both. (a) Show that the canonical map from X° to C extends to an isomorphism of a neighborhood in X with C in a neighborhood of P. (b) Show that the ramification index of $z\colon X \to S^2$ at the corresponding point is 1 in case (i), and the order of vanishing at w_0 of the function $w \mapsto F(z_0, w)$ in case (ii).

It is a theorem in analysis that every compact Riemann surface is the Riemann surface of an algebraic curve. The problem for proving this is to produce the meromorphic functions "z" and "w" on X. It is not at all obvious that there are any nonconstant meromorphic functions. For those that arise as branched coverings of S^2 (which is the same as producing one nonconstant meromorphic function z), it is not obvious how to produce another not in $\mathbb{C}(z)$; once one produces a meromorphic function w that takes n distinct values on the n points in X over some given point of S^2, however, it is not hard to see that w satisfies some irreducible equation $F(z,w) \equiv 0$, and that X is the Riemann surface of this polynomial.

Let us work out the example with $F(Z,W) = W^3 + Z^3W + Z$. By Exercise 20.2, $d(Z) = 4Z^9 + 27Z^2$, so the possible branch points are where $z = 0$, $z = \infty$, and the solutions of $z^7 = -27/4$. The points of C over the finite points are nonsingular. Over 0 there is one point $(0,0)$, and the ramification index of $z\colon X \to S^2$ is 3. Over each of the seventh roots of $-27/4$ there are two points, so one of each must have ramification index 2 and the other must be unramified. The point at infinity can be analyzed by making the substitution $Z' = 1/Z$, but one can see from the Riemann–Hurwitz formula that the sum of the numbers $e_z(P) - 1$ for the points P over ∞ must be odd, so there must be two points over ∞, with ramification indices 2 and 1. The genus g_X of the Riemann surface is given by

$$2g_X - 2 \;=\; -2n + \Sigma(e_z(P) - 1) \;=\; -6 + 2 + 7 \cdot 1 + 1\,,$$

so the genus is 3.

Exercise 20.6. For each of the following polynomials, compute a set S as in Lemma 20.3, compute the ramification indices of the points over S in the Riemann surface, and compute the genus of the Riemann surface: (i) $W^2 - \Pi_{i=1}^m(Z - a_i)$, $a_1, \ldots, a_m$ distinct complex numbers; (ii) $4W^3 - 3Z^2W + Z^3 - 2Z$; (iii) $W^3 - Z^6 + 1$; (iv) $W^3 - 3W^2 + Z^6$; and (v) $W^m + Z^m + 1$.

Exercise 20.7. (a) If a compact Riemann surface X has a meromorphic function with only one pole of order 1, show that X is isomorphic to S^2. (b) If X has a meromorphic function with two poles of order 1, or one pole of order 2, show that it gives a two-sheeted covering $X \to S^2$. (c) Show that, for a given set of $2g + 2$ points in S^2, there is, up to isomorphism, exactly one two-sheeted covering that branches at these points.

20b. Meromorphic Functions on a Riemann Surface

The meromorphic functions on a Riemann surface X form a field, which we denote by $M(X)$ or just M. If X is the Riemann surface of the polynomial $F(Z, W)$, then the functions z and w (which come from the two projections of $C \subset \mathbb{C}^2$ to the axes) are seen as in Lemma 20.4 to be meromorphic functions on X, as are any rational functions of z and w. Such rational functions form a subfield of M, which we can denote by $\mathbb{C}(z, w)$.

Proposition 20.8. *Every meromorphic function on X is a rational function of z and w:*

$$\begin{aligned} M &= M(X) = \mathbb{C}(z, w) \\ &= \mathbb{C}(z) + \mathbb{C}(z) \cdot w + \mathbb{C}(z) \cdot w^2 + \ldots + \mathbb{C}(z) \cdot w^{n-1}. \end{aligned}$$

Proof. Note first that any element in $\mathbb{C}(z)[w]$ can be written as a polynomial of degree at most $n-1$ is w, as seen by dividing by F. So it suffices to show that any meromorphic function h on X is in $\mathbb{C}(z)[w]$.

From Lemma 20.1 we have an equation

$$\begin{aligned} d(z) \cdot h &= B(z, w) \cdot F(z, w) \cdot h + C(z, w) \cdot F_w(z, w) \cdot h \\ &= C(z, w) \cdot F_w(z, w) \cdot h, \end{aligned}$$

so it suffices to show that $F_w(z, w) \cdot h$ is in $\mathbb{C}(z)[w]$. For z not in the branch set S, let $P_1, \ldots, P_n$ be the points of X over z. Then

$$\begin{aligned} F(z, W) &= a_0(z) \cdot \prod_{j=1}^{n} (W - w(P_j)), \\ F_w(z, W) &= a_0(z) \cdot \sum_{i=1}^{n} \prod_{j \neq i} (W - w(P_j)), \end{aligned}$$

so, for $1 \leq k \leq n$,

$$h(P_k) \cdot F_w(z, w(P_k)) = a_0(z) \cdot h(P_k) \cdot \prod_{j \neq k} (w(P_k) - w(P_j)).$$

Now consider the expression

$$a_0(z) \cdot \sum_{i=1}^{n} h(P_i) \cdot \prod_{j \neq i} (T - w(P_j)).$$

On the one hand, as in Lemma 20.4, this can be written in the form $\Sigma_{m=0}^{n-1} b_m(z) T^m$ with each b_m meromorphic on S^2, so $b_m \in \mathbb{C}(z)$. On the other hand, the preceding calculation shows that

$$\begin{aligned} h(P_k) \cdot F_w(z, w(P_k)) &= a_0(z) \cdot h(P_k) \cdot \prod_{j \neq k} (w(P_k) - w(P_j)) \\ &= \sum_{m=0}^{n-1} b_m(z) w(P_k)^m. \end{aligned}$$

This means that $h \cdot F_w(z, w)$ and $\Sigma b_m(z) w^m$ agree on the complement of a finite set, which implies that they are equal. □

Exercise 20.9. If $F_z(z_0, w_0) \neq 0$, show that the ramification index of $z: X \to S^2$ at the point corresponding to (z_0, w_0) is one more than the order of the meromorphic function $F_w(z, w)$ at (z_0, w_0).

With $Y = S^2$ and $f: X \to Y$ the mapping given by z, and with $p: X^\circ \to Y^\circ = Y \setminus S$ the covering space obtained by throwing away the branch points, consider the three groups:

$$\begin{aligned}
\mathrm{Aut}(X^\circ/Y^\circ) &= \{\text{continuous } \varphi: X^\circ \to X^\circ: p \circ \varphi = \varphi\};\\
\mathrm{Aut}(X/Y) &= \{\text{analytic } h: X \to X: f \circ h = h\};\\
\mathrm{Aut}(M(X)/\mathbb{C}(z)) &= \{\text{field homomorphisms } \vartheta: M(X) \to M(X): \vartheta \text{ is}\\
&\qquad \text{the identity on } \mathbb{C}(z)\}.
\end{aligned}$$

The first is topological, the second analytic, the third algebraic. We claim that they are the same, after reversing the order of multiplication in the third group:

$$\mathrm{Aut}(X^\circ/Y^\circ) \cong \mathrm{Aut}(X/Y) \cong \mathrm{Aut}(M(X)/\mathbb{C}(z))^{\mathrm{opp}}.$$

Exercise 20.10. Prove this by showing that every deck transformation φ extends uniquely to an analytic isomorphism h, and showing that every automorphism of $M(X)$ that is the identity on $\mathbb{C}(z)$ has the form $f \mapsto f \circ h$ for a unique h. If you know some Galois theory, show that $M(X)$ is a Galois extension of $\mathbb{C}(z)$ if and only if the covering X° of Y° is a regular covering. Show that the Riemann surfaces of two algebraic curves are isomorphic if and only if their fields of meromorphic functions are isomorphic $\mathbb{C}$-algebras. Show that for any finite group G there is a Galois extension L of $\mathbb{C}(z)$ whose Galois group is G.

In fact, for a given compact Riemann surface Y, to give a compact Riemann surface X with a nonconstant analytic map from X to Y is equivalent to giving a finite-sheeted topological covering $X^\circ \to Y^\circ$ of the complement of a finite set, or to specify a finite extension $M(X)$ of the field $M(Y)$ of meromorphic functions. In this setting, the similarity seen in Proposition 13.23 between coverings and field extensions is more than just an analogy.

In most expositions the Riemann surface of a polynomial is constructed, following Weierstrass, by starting with a germ of an analytic function $w(z)$ satisfying the equation $F(z, w(z)) \equiv 0$, and analytically continuing it around the plane. This approach, however, does not take advantage of the fact that the algebraic curve C is already, except for the modification and addition of a finite number of points, the desired Riemann surface.

Problem 20.11. Carry out this construction, using the ideas of §16b, and show that it gives the same Riemann surface.

20c. Holomorphic and Meromorphic 1-Forms

For any differentiable surface, as in Chapter 9, we can consider not just real differentiable 1-forms but complex ones as well, where a complex 1-form is given by $\omega_1 + i\omega_2$, with ω_1 and ω_2 real 1-forms. Again we can consider closed and exact forms, with $\omega_1 + i\omega_2$ being closed (resp. exact) when each of ω_1 and ω_2 is closed (resp. exact). The corresponding group of closed complex 1-forms modulo exact complex 1-forms is denoted $H^1(X;\mathbb{C})$. It is a complex vector space, which can be identified with $H^1(X) \oplus iH^1(X)$. If X is compact of genus g_X, $H^1(X;\mathbb{C})$ is a complex vector space of dimension $2g_X$.

When X is a compact Riemann surface, and not just a differentiable surface, there are some special closed complex 1-forms, called *holomorphic* 1-forms. They are the 1-forms that in local coordinates $\varphi_\alpha \colon U_\alpha \to X$, with $U_\alpha \subset \mathbb{C}$, have the form

$$f_\alpha\, dz = (u_\alpha + iv_\alpha)(dx + i\, dy) = (u_\alpha\, dx - v_\alpha\, dy) + i(v_\alpha\, dx + u_\alpha\, dy),$$

where $f_\alpha(z) = f_\alpha(x+iy) = u_\alpha(x,y) + iv_\alpha(x,y)$ is an analytic function. To define a global 1-form, using the notation of §19a, we must have

$$f_\alpha(z) = f_\beta(\varphi_{\beta\alpha}(z)) \cdot \varphi_{\beta\alpha}{}'(z) \quad \text{on } U_{\beta\alpha}, \qquad \text{where } \varphi_{\beta\alpha}{}' = \frac{d\varphi_{\beta\alpha}}{dz}$$

is the complex derivative. As before, the Cauchy–Riemann equations $\partial u_\alpha/\partial x = \partial v_\alpha/\partial y$ and $\partial u_\alpha/\partial y = -\partial v_\alpha/\partial x$ say that such a 1-form is closed.

The holomorphic 1-forms form a complex vector space, sometimes denoted $\Omega^{1,0}(X)$, or just $\Omega^{1,0}$. A holomorphic 1-form ω is exact precisely when there is an analytic function g on X with $dg = \omega$. In particular, if X is compact, every analytic function is constant, and so if ω is exact, then $\omega = 0$. This means that the natural map from $\Omega^{1,0}$ to $H^1(X;\mathbb{C})$ is injective. We regard $\Omega^{1,0}$ as a complex subspace of $H^1(X;\mathbb{C})$.

There is also a notion of an *antiholomorphic* 1-form. This is a 1-form that locally has the form

$$\overline{f_\alpha}\,\overline{dz} \;=\; (u_\alpha - iv_\alpha)(dx - i\,dy) = (u_\alpha\,dx - v_\alpha\,dy) + i(-v_\alpha\,dx - u_\alpha\,dy)$$

with $f_\alpha = u_\alpha + iv_\alpha$ an analytic function as above. Again these form a complex vector space, denoted $\Omega^{0,1}(X)$ or $\Omega^{0,1}$, and again these are closed forms, and the only antiholomorphic forms which are exact are differentials of complex conjugates $\overline{g}$ of complex analytic functions. Again, for X compact, we regard $\Omega^{0,1}$ as a subspace of $H^1(X;\mathbb{C})$.

It follows readily from the definitions that no nonzero 1-form can be both holomorphic and antiholomorphic: $\Omega^{1,0} \cap \Omega^{0,1} = 0$. For any surface X the space $H^1(X;\mathbb{C})$ has a *complex conjugation* operator, that takes $\omega = \omega_1 + i\omega_2$ to $\overline{\omega} = \omega_1 - i\omega_2$. This is not a complex linear map, but is *conjugate linear*: it is linear as a map of real spaces, and for a complex number c, $\overline{c\omega} = \overline{c\omega}$. Complex conjugation takes $\Omega^{1,0}$ to $\Omega^{0,1}$ and $\Omega^{0,1}$ to $\Omega^{1,0}$. The following is one of the major "existence theorems" about compact Riemann surfaces:

Theorem. *For a compact Riemann surface X, $H^1(X;\mathbb{C}) = \Omega^{1,0} \oplus \Omega^{0,1}$. Equivalently, $dim(\Omega^{1,0}) = dim(\Omega^{0,1}) = g_X$.*

In light of the preceding paragraph, the theorem is equivalent to showing that there are g_X linearly independent holomorphic 1-forms on X. We will say nothing about the proof of this theorem for a general compact Riemann surface, except to say that producing such 1-forms is closely related to producing meromorphic functions, which we have already discussed. For example, if ω_1 and ω_2 are two independent holomorphic 1-forms, then $\omega_2 = f \cdot \omega_1$, where f is a nonconstant meromorphic function. When the Riemann surface comes from an algebraic curve, however, we will prove this theorem in the next chapter. If the curve has a sufficiently nice form, however, it can be proved directly, as in the following problems:

Problem 20.12. Suppose $F(Z, W) = \Sigma_{i=0}^{n} a_i(Z)W^{n-i}$ is the polynomial as in §20a, and assume: (i) the curve C is nonsingular, i.e., there are no points (z_0, w_0) at which F, F_w, and F_z all vanish; and (ii) the degree of $a_i(Z)$ as a polynomial in Z is at most i, and if λ_i is the coefficient of Z^i in a_i, the equation $\Sigma_{i=0}^{n} \lambda_i t^{n-i} = 0$ has n distinct roots. (a) Show that X has n points over $\infty \in S^2$. (b) Show that $h = F_w(z, w)$ has a pole of order $n-1$ at each of these n points. (c) Show that the genus of

X is $(n-1)(n-2)/2$. (d) Show that

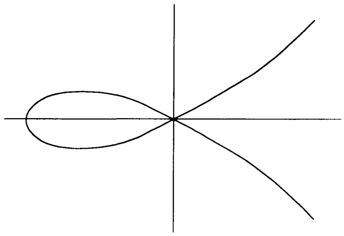
$$\Omega^{1,0}(X) = \left\{ \frac{g(z,w)}{F_w(z,w)} dz \colon g \text{ is a polynomial of degree at most } n-3 \text{ in } z \text{ and } w \right\},$$

which is a complex vector space of dimension g_X.

The following problem generalizes this to allow the curve C to have the simplest possible singularities: simple *nodes*. A node is a point where the two partial derivatives vanish, but the Hessian

$$\left(\frac{\partial^2 F}{\partial z\, \partial w}\right)^2 - \frac{\partial^2 F}{\partial z^2} \cdot \frac{\partial^2 F}{\partial w^2}$$

is not zero. Geometrically, this means the curve C has two nonsingular branches that cross transversely; in particular, there are two points of the Riemann surface over a node of C. The corresponding picture of a real curve ($W^2 = Z^2 + Z^3$) is:



Problem 20.13. Generalize the preceding problem to allow C to have a certain number δ of nodes, but no other singularities, but continue to assume (ii). Show that the genus of X is $(n-1)(n-2)/2 - \delta$, and that the holomorphic 1-forms on X are exactly those that have the form $(g(z,w)/F_w(z,w))\,dz$ where g is a polynomial of degree at most $n-3$ in z and w that vanishes at the nodes of C.

It is an important fact from the theory of algebraic curves that, after suitable algebraic transformations, every algebraic curve can be put in the form considered in the preceding problem. Note that it is necessary to allow singularities, since not every genus has the form $(n-1)(n-2)/2$. For a nice discussion of this, see Griffiths (1989).

One can also consider *meromorphic* 1-forms on X, which can be defined as for holomorphic 1-forms, locally given by $f_\alpha dz$, but with f_α only required to be meromorphic. If ω is a meromorphic 1-form, the order of ω at a point P, denoted $\mathrm{ord}_P(\omega)$, is defined to be the order of vanishing of f_α at the corresponding point in a coordinate disk. For example, when $X = S^2$, and $\omega = dz = -(z')^{-2}dz'$, with $z' = 1/z$, then ω has a pole of order 2 at the point at infinity. In general, we have

Proposition 20.14. *For any nonzero meromorphic* 1*-form ω on a compact Riemann surface X,*

$$\sum_{P\in X} \mathrm{ord}_P(\omega) \;=\; 2g_X - 2.$$

Proof. It is enough to prove the formula for one such ω, since any other has the form $h\cdot\omega$ for some meromorphic function h, and

$$\sum_{P\in X} \mathrm{ord}_P(h\cdot\omega) \;=\; \sum_{P\in X} \mathrm{ord}_P(h) + \mathrm{ord}_P(\omega) \;=\; 0 + \sum_{P\in X} \mathrm{ord}_P(\omega)$$

by Corollary 19.5. Assume that X comes equipped with a nonconstant meromorphic function $f\colon X \to S^2$. We take $\omega = df = f^*(dz)$. Near a point P of X, where the mapping in local coordinates is $t \mapsto t^e$, a meromorphic form on S^2 with local expression $g(z)\,dz$ pulls back to one on X with local expression $g(t^e)et^{e-1}\,dt$. Therefore $\mathrm{ord}_P(\omega) = (e_f(P)-1) + e_f(P)\,\mathrm{ord}_{f(P)}(dz)$, and since the sum of $e_f(P)$ for P mapping to the point at infinity is n, and $\mathrm{ord}_\infty(dz) = -2$,

$$\sum_{P\in X} \mathrm{ord}_P(\omega) \;=\; \sum_{P\in X} (e_f(P)-1) + n(-2);$$

this is $2g_X - 2$ by the Riemann–Hurwitz formula. □

Another proof can be given by appealing to the result of Chapter 8. Given a meromorphic 1-form ω with local expression $f_\alpha dz$, define a vector field V whose expression in the same local coordinates is given by $V_\alpha = 1/f_\alpha$, i.e., if $f_\alpha = u_\alpha + iv_\alpha$, then

$$V_\alpha \;=\; \left(\frac{u_\alpha}{u_\alpha^{\,2}+v_\alpha^{\,2}},\, \frac{-v_\alpha}{u_\alpha^{\,2}+v_\alpha^{\,2}}\right).$$

Exercise 20.15. Verify that these V_α define a vector field V on X, and that $\mathrm{Index}_P V = -\mathrm{ord}_P(\omega)$.

Proposition 20.14 therefore also follows from Theorem 8.3.

Exercise 20.16. Reverse the argument in the above proof to give another proof of the Riemann–Hurwitz formula for $f\colon X \to Y$, under the assumption that Y has a nonzero meromorphic 1-form.

Define the *residue* of a meromorphic 1-form ω at a point P on a Riemann surface X, denoted $\mathrm{Res}_P\,\omega$, to be $(1/2\pi i)\int_\gamma \omega$, where γ is a small counterclockwise circle around P not surrounding any point except P where ω is not holomorphic.

Exercise 20.17. (a) Show that this is well defined. (b) If z is a local coordinate at P, and $\omega = f(z)\,dz$, with $f(z) = \Sigma_{n=-m}^{\infty} a_n z^n$, show that $\mathrm{Res}_P(\omega) = a_{-1}$.

Proposition 20.18 (Residue Formula). *If ω is a meromorphic* 1*-form on a compact Riemann surface X, then*

$$\sum_{P\in X} \mathrm{Res}_P(\omega) \;=\; 0\,.$$

Proof. We know that we can realize X as a polygon Π with sides identified. From the construction we see that the map from Π to X can be taken to be differentiable, and, moving the sides slightly, we may assume the image of each side is disjoint from the set of poles of ω. The 1-form ω then determines a closed 1-form ω' on Π, and we must show that the sum of the integrals of ω' around small circles around the poles of ω' is zero. Corollary 9.12 implies that this sum is the same as the integral of ω' around the boundary of Π, and this integral vanishes since the integrals over the sides that get identified in X cancel in pairs. □

The following problem gives another proof for Riemann surfaces coming from algebraic curves:

Problem 20.19. (a) Prove the Residue Formula directly when $X = S^2$. (b) With $z\colon X \to S^2$ as in §20b, and $\omega = f\,dz$, with $f \in M(X)$, show that, for $Q \in S^2$,

$$\sum_{P\in z^{-1}(Q)} \mathrm{Res}_P(f\,dz) \;=\; \mathrm{Res}_Q(g\,dz)\,,$$

where g in $\mathbb{C}(z)$ is the trace of the $\mathbb{C}(z)$-linear endomorphism of $M(X)$ that is left multiplication by f. (c) Deduce the Residue Formula for $\omega = f\,dz$ from (a) and (b).

20d. Riemann's Bilinear Relations and the Jacobian

We have seen that the space $\Omega^{1,0}(X)$ of holomorphic 1-forms on X is a subspace of the De Rham group $H^1(X;\mathbb{C})$. The way $\Omega^{1,0}(X)$ sits in $H^1(X;\mathbb{C})$ is important both in studying functions on X and in studying the moduli of Riemann surfaces of given genus. Here we indicate how some of this is related to the facts about homology that we proved in Chapter 18. For this we use the pairing $(\omega, \nu) = \iint_X \omega \wedge \nu$ defined on closed 1-forms and their cohomology classes, but extended linearly to those with complex coefficients as usual. When applied to holomorphic 1-forms, there are two simple consequences of the definition:

(i) $(\omega, \nu) = 0$ if ω and ν are holomorphic; and
(ii) $i \cdot (\omega, \overline{\omega}) > 0$ if ω is nonzero and holomorphic.

The first follows from the fact that $dz \wedge dz = 0$. For the second, if, in local coordinates, $\omega = f(z)\, dz = f(z)(dx + i\,dy)$, then

$$i \cdot \omega \wedge \overline{\omega} \;=\; i \cdot |f(z)|^2 (dx + i\,dy) \wedge (dx - i\,dy) \;=\; 2|f(z)|^2\, dx \wedge dy,$$

which is strictly positive wherever f is not zero, so its integral is positive.

Taking a basis $a_1, \ldots, a_g, b_1, \ldots, b_g$ for homology as in Chapter 18, and applying Exercise 18.8, we deduce, for ω and ν holomorphic as above:

Proposition 20.20 (Riemann's Bilinear Relations).

(1) $\displaystyle\sum_{j=1}^{g} \int_{a_j} \omega \int_{b_j} \nu \;=\; \sum_{j=1}^{g} \int_{a_j} \nu \int_{b_j} \omega;$

(2) $\displaystyle i \cdot \sum_{j=1}^{g} \left(\int_{a_j} \omega \int_{b_j} \overline{\omega} - \int_{a_j} \overline{\omega} \int_{b_j} \omega \right) \;>\; 0.$

Corollary 20.21. *There is a unique basis $\omega_1, \ldots, \omega_g$ for the space of holomorphic 1-forms so that*

$$\int_{a_j} \omega_k \;=\; \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{otherwise}. \end{cases}$$

Proof. This follows immediately from the fact that $H^1(X;\mathbb{C})$ is the direct sum of $\Omega^{1,0}(X)$ and $\Omega^{0,1}(X)$, together with the fact that integrating over cycles gives an isomorphism $H^1(X;\mathbb{C}) \cong \operatorname{Hom}(H_1X, \mathbb{C})$. $\square$

The integrals of this basis over the other basis elements $b_1, \ldots, b_g$ carry important information. Let

$$\tau_{j,k} = \int_{b_k} \omega_j, \qquad 1 \leq j, k \leq g.$$

This gives a $(g \times g)$-matrix of complex numbers $Z = (\tau_{j,k})$, called the (normalized) *period matrix* corresponding to the homology basis. This matrix Z is far from arbitrary. It follows from what we have just seen that Z is nonsingular. Moreover,

Corollary 20.22. *(i) The matrix Z is symmetric, i.e., $\tau_{k,j} = \tau_{j,k}$. (ii) The matrix $Im(Z)$ whose entries are the imaginary parts of the entries of Z is a positive-definite symmetric matrix.*

Proof. (i) is an immediate consequence of the first of Riemann's bilinear relations, applied to the forms ω_j and ω_k.

(ii) follows from the second, applied to a form $\omega = \Sigma_{j=1}^{g} t_j \omega_j$, with t_j arbitrary real numbers, not all zero:

$$0 < i \cdot (\omega, \overline{\omega}) = i \cdot \sum_{j,k} t_j t_k (\overline{\tau}_{k,j} - \tau_{j,k}) = 2 \sum_{j,k} t_j t_k (\mathrm{Im}(\tau_{j,k}). \qquad \square$$

Of course, the period matrix is not unique, depending as it does on the choice of a homology basis, but if another basis is chosen, with the same intersection numbers, they will differ by a nonsingular matrix with integral entries that preserves this intersection pairing.

As in the plane, integrals between two points on X are determined up to these periods. Periods also play a role in a fundamental theorem of Abel, which concerns the question of when one can find a meromorphic function on a given Riemann surface with zeros and poles of given orders at given points.

Exercise 20.23. Show that for $X = S^2$, for any finite set of points P_i and any given integers m_i, provided $\Sigma m_i = 0$, there is a meromorphic function f on X with $\mathrm{ord}_{P_i}(f) = m_i$ for all i.

For $g > 0$, taking any basis $\omega_1, \ldots, \omega_g$ of the holomorphic 1-forms, we have a mapping from the homology group H_1X to $\mathbb{C}^g$ given by

$$\gamma \mapsto \left(\int_\gamma \omega_1, \ldots, \int_\gamma \omega_g\right).$$

It follows from what we have seen above that this map embeds H_1X in $\mathbb{C}^g$ as a *lattice*, i.e., the image of the basis elements $a_1, \ldots, b_g$

form a basis for $\mathbb{C}^g$ as a real vector space. If this lattice is denoted by Λ, then the quotient space (and group) $\mathbb{C}^g/\Lambda$ is called the *Jacobian* of X, and denoted $J(X)$; it is homeomorphic to $\mathbb{R}^{2g}/\mathbb{Z}^{2g}$, so to the Cartesian product of $2g$ circles.

If P and Q are any two points in X, they determine a point denoted $[Q-P]$ in $J(X)$, by the formula

$$[Q-P] = \left(\int_P^Q \omega_1, \ldots, \int_P^Q \omega_g\right) \in \mathbb{C}^g/\Lambda = J(X),$$

where the notation means to integrate along any path from P to Q; the resulting vector is defined up to an element in Λ. Similarly, given any 0-cycle $D = \Sigma m_i P_i$ of degree zero on X, one can define a point $[D]$ in the Jacobian by writing $D = \Sigma(Q_j - P_j)$ and setting $[D]$ equal to $\Sigma[Q_j - P_j]$. Equivalently, fix a point P_0 in X, and define the Abel–Jacobi mapping $A\colon X \to J(X)$ by the formula $A(P) = [P - P_0]$. Then $[\Sigma m_i P_i] = \Sigma m_i A(P_i)$.

Exercise 20.24. Show that this gives a well-defined homomorphism from the group $\tilde{Z}_0 X$ of 0-cycles of degree zero on X to $J(X)$.

The map from H_1X to $\mathbb{C}^g$ can be defined intrinsically, without choosing a basis of holomorphic 1-forms, as the map

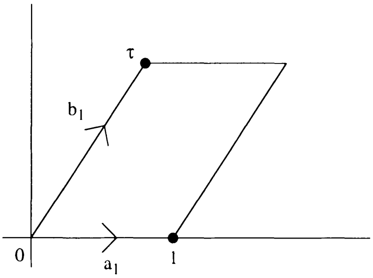
$$H_1X \to \Omega^{1,0}(X)^*, \qquad \gamma \mapsto \left[\omega \mapsto \int_\gamma \omega\right],$$

where $\Omega^{1,0}(X)^*$ is the dual space of complex-valued functions on $\Omega^{1,0}(X)$. This gives $J(X) = \Omega^{1,0}(X)^*/H_1(X)$, without choices.

The *divisor* $\mathrm{Div}(f)$ of a meromorphic function f is the 0-cycle $\Sigma\, \mathrm{ord}_P(f)P$. In the next chapter we will prove *Abel's theorem* that a zero cycle $D = \Sigma m_i P_i$ is the divisor of a meromorphic function if and only if its degree Σm_i is zero and $[D]$ is zero in $J(X)$. Equivalently, D is the boundary of a 1-chain γ such that $\int_\gamma \omega = 0$ for all holomorphic 1-forms ω. We will also prove the *Jacobi inversion theorem* that the Abel–Jacobi map from $\tilde{Z}_0 X$ to $J(X)$ is surjective.

20e. Elliptic and Hyperelliptic Curves

Before turning to the general situation it may be helpful to work out the simplest nontrivial case in detail. Consider a Riemann surface of the form $\mathbb{C}/\Lambda$, where Λ is a lattice. By Exercise 19.17 we may choose such a lattice to be generated by 1 and τ, where τ is in the upper half plane. Realizing X by identifying the sides of the parallelogram



the images of the indicated sides can be taken as the standard basis a_1 and b_1 for H_1X. Now X has a natural holomorphic 1-form ω, whose pull-back to $\mathbb{C}$ is just the 1-form dz, with z the standard coordinate on $\mathbb{C}$. The integral of ω along a_1 is clearly 1, and the integral of ω along b_1 is τ. (Note that the corresponding period matrix is $Z = (\tau)$, in agreement with Corollary 20.22.) The mapping from H_1X to $\mathbb{C} = \mathbb{C}^g$, $\gamma \mapsto \int_\gamma \omega$, has image Λ generated by 1 and τ. Fix the point P_0 in X to be the image of the origin in $\mathbb{C}$. The Abel–Jacobi mapping from X to $\mathbb{C}/\Lambda$ then takes a point P to $[P - P_0] = \int_{P_0}^{P} \omega$. By looking at this mapping on the parallelogram spanned by 1 and τ, we see that this mapping is an isomorphism, i.e., $A\colon X \xrightarrow{\cong} \mathbb{C}/\Lambda$.

Now suppose X is any compact Riemann surface of genus 1, which we assume to have a nonzero holomorphic 1-form ω. We have the mapping $H_1X \to \mathbb{C}$ taking a homology class γ to $\int_\gamma \omega$. The image is a lattice Λ in $\mathbb{C}$. If we take a basis a_1, b_1 for H_1X as usual, we can take ω so that $\int_{a_1} \omega = 1$, in which case Λ is generated by 1 and $\tau = \int_{b_1} \omega$. As a very special case of Corollary 20.22 it follows that τ is in the upper half-plane. Fix a point P_0 in X.

Proposition 20.25. *The Abel–Jacobi mapping*

$$X \to \mathbb{C}/\Lambda, \qquad P \mapsto [P - P_0] = \int_{P_0}^{P} \omega,$$

is an isomorphism of X with $\mathbb{C}/\Lambda$.

Proof. Since the Abel–Jacobi mapping is analytic and nonconstant, it is unramified by the Riemann–Hurwitz formula, and therefore a covering map. We know from Chapter 13 that all finite coverings of $\mathbb{C}/\Lambda$ are given by subgroups of the fundamental groups, which means that X has the form $\mathbb{C}/\Lambda'$, where $\Lambda' \subset \Lambda$ is a subgroup of finite index.

But we proved directly that for any X of this form, the Abel–Jacobi map is an isomorphism. □

If a Riemann surface X is a two-sheeted covering of S^2 with four branch points, then by the Riemann–Hurwitz formula it has genus 1. By changing the map by an automorphism of S^2 one can take these four branch points to be 0, 1, ∞, and another $\lambda \in \mathbb{C}$. It follows easily (see below for a more general situation) that X is the Riemann surface of the curve $W^2 = Z(Z-1)(Z-\lambda)$. Since dz/w is a holomorphic 1-form on this Riemann surface, we know from what we have just seen that if we fix P_0 on X, the map

$$P \mapsto \int_{P_0}^{P} \frac{dz}{w} = \int_{P_0}^{P} \frac{dz}{\sqrt{z(z-1)(z-\lambda)}}$$

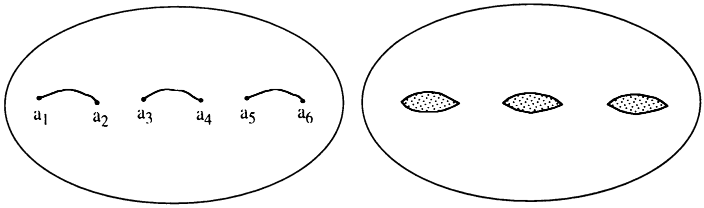
gives an isomorphism from X to $\mathbb{C}/\Lambda$. We will see that every Riemann surface of genus 1 is isomorphic to those arising this way, at least under the assumption that X comes from an algebraic curve. Most complex analysis texts have a chapter on such "elliptic integrals." (The Weierstrass $\wp$-function is used to find a two-sheeted branched covering from $\mathbb{C}/\Lambda$ to S^2.)

We end this brief excursion by looking at the special case of Riemann surfaces that can be realized as two-sheeted branched coverings of the sphere. By the Riemann–Hurwitz formula, such a covering must have an even number of branch points, namely, $2g_X + 2$. By Exercise 20.7, there is only one two-sheeted covering of S^2 with a given set of an even number of branch points. If these are the points $a_1, \ldots, a_m$ in $\mathbb{C}$, possibly together with the point ∞ (if m is odd), this can be realized as the Riemann surface of the algebraic curve

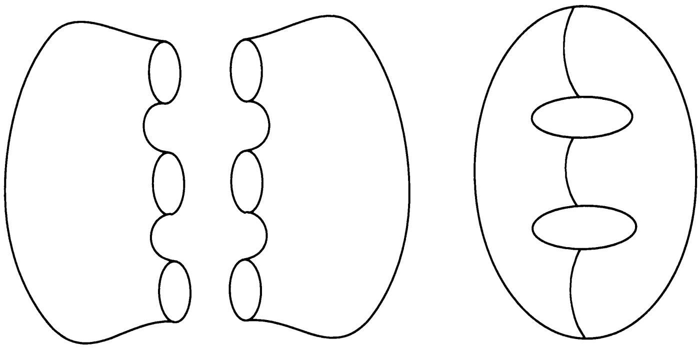
$$W^2 - \prod_{i=1}^{m} (Z - a_i) = 0.$$

These curves, and the corresponding Riemann surfaces, are called *hyperelliptic*. All Riemann surfaces of genus 2 are in fact hyperelliptic, but for genus greater than 2, not all Riemann surfaces arise this way.

The topology of a hyperelliptic surface can be seen directly, by cutting slits in the sphere along arcs from a_1 to a_2, a_3 to $a_4, \ldots, a_{2g+1}$ to a_{2g+2}.



The two-sheeted covering over the complement of these slits is trivial, so the Riemann surface can be constructed as in the picture:



Exercise 20.26. If $m = 2g_X + 2$, verify that the 1-forms $z^i\,dz/w$, for $0 \le i \le g-1$, are holomorphic, and therefore give a basis for the holomorphic 1-forms.